The book develops six major layers of the subject, each building on the last:

The opening part constructs the coordinate-free language of smooth manifolds, tensors, differential forms, and Stokes' theorem from first principles, using the Schwarzschild horizon as a running motivation throughout. Curvature is developed as a geometric object, not a transformation rule, and the distinction between genuine spacetime singularities and coordinate artifacts is made precise and operational.

The second part develops the four classical families of exact black hole solutions: Schwarzschild, Kruskal, Kerr, and Kerr–Newman. Penrose conformal diagrams are derived carefully and used to read off causal structure at a glance. The Kerr solution's hidden Carter constant, the physics of the ergosphere, frame dragging, and the Penrose process are all treated with full geometric precision.

The third part, one of the most distinctive in the book, develops horizon geometry as a field of study in its own right. Event horizons, apparent horizons, trapping horizons, isolated horizons, and dynamical horizons are carefully distinguished, because different horizon concepts satisfy different theorems and obey different thermodynamic laws. The Raychaudhuri equation and the focusing of null congruences are derived and used to prove both the Penrose singularity theorem and the classical black hole area increase theorem.

The fourth part equips the reader with the tools of modern black hole research: the 3+1 decomposition that underlies numerical relativity, the Newman–Penrose formalism and its Weyl scalars, black hole perturbation theory and the Regge–Wheeler and Teukolsky equations, quasinormal modes, and the optical geometry of black hole shadows and gravitational lensing.

The fifth part develops black hole thermodynamics geometrically. The Hawking temperature is derived from the requirement that the Euclidean continuation of the black hole geometry be smooth, an argument that requires no quantum field theory and reveals the result as a consequence of pure geometry. The Bekenstein–Hawking entropy formula is derived through the Wald Noether charge method and connected to holography and the Ryu–Takayanagi formula for entanglement entropy in anti-de Sitter spacetime.

The final part surveys the frontiers: higher-dimensional black holes and the Gregory-Laflamme instability, the Penrose and Hawking singularity theorems in full geometric form, modified gravity and non-Riemannian geometry, and numerical relativity as computational differential geometry.



S M Nazmuz Sakib is an interdisciplinary scholar, engineer, and author bridging IT, business analytics, and applied research. He is a Fellow of the Scholars Academic and Scientific Society, a Member of IAENG, a lifetime member of BELTA, and a Bangladeshi Arena Grand Master (AGM) recognized by FIDE. His credentials include extensive certifications across AWS, Google/SAS analytics, UX, project management, digital marketing, and finance. Dr. Sakib has published several peer-reviewed articles and book chapters with Springer Nature, CRC Press/Taylor & Francis, and IGI Global on various topics.

Black Holes Geometry S M Nazmuz Sakib

GRADUATE TEXTS IN MATHEMATICAL PHYSICS

Black Holes
Geometry

CURVATURE · HORIZONS · SPACETIME

DIFFERENTIAL GEOMETRY OF EXTREME GRAVITY

PART I – III
Manifolds, Tensors & Horizon Geometry

PART IV – V
Perturbations, Thermodynamics & Holography

PART VI
Singularity Theorems & Numerical Relativity

Black Holes Geometry

Curvature and Horizons

S M Nazmuz Sakib

S M Nazmuz Sakib

Black Holes Geometry: Curvature and Horizons

S M Nazmuz Sakib

Chapter	Title	Page
Part I: Geometric Foundations		**22**
1	Manifolds, Tensors, and Physics as Coordinate Free Thought	22
2	Connections, Curvature, and Parallel Transport	46
3	Lorentzian Geometry: Time Joins the Chat	70
4	Symmetries as Geometry: Killing Fields and Conserved Quantities	95
Part II: Einstein Geometry and Exact Black Hole Spacetimes		**121**
5	Einstein's Equations as Curvature Constraints	121
6	Schwarzschild Geometry: The Simplest Nontrivial Monster	151
7	Global Structure and Penrose Diagrams (Causality as Artwork)	174
8	Kerr Geometry: Rotation Warps Reality in a New Dimension	197
9	Charged and Exotic Variants: Kerr-Newman and Friends	221
Part III: Horizon Geometry as a Field of Study		**247**
10	Horizons: Event, Apparent, Trapping, Isolated, Dynamical	247
11	Null Congruences and Raychaudhuri: Focusing as Theorem Machine	270

12	Quasi-local Mass and Geometric Inequalities	297
Part IV: Geometric Methods for Dynamics and Perturbations		**324**
13	3+1 Geometry and the Initial Value Problem	324
14	Tetrads, Spinors, and Newman-Penrose Geometry	351
15	Perturbations of Black Hole Geometry	375
16	Optical Geometry, Lensing, and Finsler Randers Tricks	403
Part V: Thermodynamics, Euclidean Methods, and Quantum Geometry		**430**
17	Euclidean Black Hole Geometry and Temperature	430
18	Entropy, Area, and Geometric Microstories	461
19	Holography and Hyperbolic Geometry: AdS as a Geometry Playground	482
Part VI: Extensions and Frontiers		**505**
20	Higher-Dimensional Black Hole Geometries	505

21	Topology, Censorship, and Singularity Theorems	539
22	Modified Gravity and Non-Riemannian Geometry	563
23	Numerical Relativity as Computational Differential Geometry	575

Black Holes Geometry: Curvature and Horizons

S M Nazmuz Sakib

Preface

Black Holes Geometry: Curvature and Horizons

The Geometry That Cannot Be Faked

Black holes have a peculiar way of testing the honesty of the mathematics used to describe them. In most of classical physics, a theorist can choose convenient coordinates, carry out a calculation in that preferred frame, and trust that the final numbers will survive the journey back into physical reality. Black hole physics does not extend that trust for long. The moment a horizon appears on the stage, the coordinates that served perfectly well everywhere else begin to misbehave. Perfectly regular geometric points acquire the appearance of catastrophe. Smooth, traversable surfaces look singular in one labeling scheme and entirely ordinary in another. An observer falling freely through an event horizon feels nothing unusual at the moment of crossing, yet the coordinates of a distant stationary observer suggest that the falling observer never arrives at all, hovering indefinitely at the horizon while being stretched across an infinite range of coordinate time and fading into an ever-deepening gravitational redshift.

None of this is real. It is an artifact of the map, not a feature of the territory.

This is not a minor technical inconvenience, and it is not a quirk that can be patched over with a footnote. It is the discipline's central diagnostic: any mathematical framework that cannot reliably distinguish genuine physical singularities from coordinate artifacts is, at its foundation, untrustworthy. The only known cure is to learn to think geometrically, in the precise mathematical sense that the word demands. Tensors must be understood as objects that carry their meaning intact through any change of labeling scheme, not as arrays of numbers that happen to transform in a particular way. Horizons must be recognized not as places where some metric coefficient misbehaves, but as causally defined null hypersurfaces with a rich and well-characterized intrinsic geometry of their own. Curvature must be read from invariant scalars that do not depend on how one has chosen to orient the axes, not from components that depend entirely on that choice.

The geometry comes first. Coordinates come second, as a convenient representation of the geometry, never as a substitute for it.

This book is a systematic attempt to develop that geometric fluency and to carry it from the foundational definitions all the way through to the frontiers of what is actively being computed and observed today.

Why Another Book on Black Holes

The question is worth answering directly, because the literature is genuinely rich. There are excellent introductory texts that build physical intuition for rotating and non-rotating black holes through accessible narrative and carefully chosen examples. There are advanced mathematical monographs that develop the geometry of Lorentzian manifolds with full rigor. There are specialized references for numerical relativity, for the thermodynamics of black holes, for gravitational wave source modeling, and for the holographic correspondence between gravity and field theory.

What is considerably harder to find is a treatment that moves coherently from the geometric foundations all the way to the active research frontiers, maintaining mathematical precision throughout while keeping one eye on real observations and the other on the conceptual architecture that makes the subject hang together as a unified whole.

That gap matters now more than it did even twenty years ago. Modern black hole research is data-rich in a way it has never previously been. Gravitational wave observatories have detected and characterized dozens of binary black hole mergers, generating catalogs of inferred parameters that allow population-level statistical analysis for the first time. The Event Horizon Telescope collaboration has produced resolved images of the shadow cast by the supermassive black hole at the center of the galaxy M87 and, subsequently, by the black hole at the center of our own Milky Way. Numerical relativity codes now routinely simulate the full inspiral, merger, and ringdown of binary black hole systems in regimes where no analytical solution exists. X-ray observatories and radio interferometry arrays provide additional windows into accretion flows and jets in strong-field black hole environments.

Making productive contact with all of this output requires not merely familiarity with the equations of general relativity but a working command of the geometric language in which those equations are most naturally expressed. A student who has learned to manipulate indices but not to think in terms of coordinate-free objects will find that language opaque when they encounter horizon-tracking algorithms in numerical codes, quasi-local mass definitions in waveform analysis pipelines, or the geometric reasoning that underlies the derivation of the Hawking temperature. A researcher comfortable with the formal definitions but unaccustomed to connecting them to observational quantities will find the data literature similarly opaque. This book is designed to build both capacities simultaneously, treating the abstract geometry and the concrete physics as two aspects of a single subject rather than as parallel disciplines that happen to share vocabulary.

The Central Argument: Coordinates Are a Polite Lie

This phrase appears in Chapter 1, and it is worth explaining at length here in the preface because it sets the intellectual tone for everything that follows.

A coordinate chart on a manifold is like a street map of a city. The map is genuinely useful. Without it, navigation in an unfamiliar place is nearly impossible, and no reasonable person would refuse to use one on abstract philosophical grounds. But the streets of the city do not

actually run in the directions printed on the page. The curvature of the Earth is suppressed. Distances near the edges of the map are distorted relative to distances at the center. The map is a locally faithful representation of something whose global structure it cannot capture in a single image, and it breaks down or becomes ambiguous at the boundaries of its range of validity.

Coordinate charts on spacetime work the same way. The classic coordinates used to describe a non-rotating black hole are an excellent tool for everything happening far from the horizon: computing orbital periods, predicting gravitational redshifts, calculating the bending of light around the Sun, matching to solar system observations. But at the Schwarzschild radius, where the horizon sits, these coordinates produce metric components that diverge or vanish. A naive reading would conclude that something physically catastrophic happens at that surface. Nothing does. The curvature of spacetime at that point, measured by genuine scalars constructed from the curvature tensor in ways that do not depend on any coordinate choice, is finite and unremarkable. A freely falling observer notices nothing special at the crossing. The terrible behavior is in the chart, not in the manifold.

When one passes to coordinates that cover the entire black hole spacetime without developing this kind of degeneracy, the horizon becomes a pair of null surfaces crossing at a distinguished two-sphere. A white hole region and a second asymptotically flat exterior, completely invisible in the original coordinates, emerge as natural geometric features of the same solution. These are not new physical objects invented by a change of variables. They were always there. The original chart was concealing them because its time coordinate breaks down at the horizon in a way that has nothing to do with the spacetime itself.

The same lesson appears throughout the subject. The coordinates most commonly used for rotating black holes have their own degeneracy at the outer horizon, concealing what is actually a smooth geometric surface. The inner horizon of a charged black hole is invisible in exterior coordinates. The ring singularity of a rotating black hole and the full global structure of its spacetime require a careful collection of overlapping coordinate patches to describe correctly, none of which is individually adequate. The Newman-Penrose formalism, tetrad methods, Penrose conformal diagrams, and the splitting of spacetime into space plus time that underlies numerical relativity are all, in different ways, tools for extracting the coordinate-independent geometric content from spacetimes that any single coordinate system describes inadequately.

The coordinate-free framework developed in this book is not a philosophical luxury for purists. It is a form of error control. It is how one avoids being misled by the map into believing something false about the territory.

The Architecture of the Book

The book is organized into six parts, each addressing a distinct and progressively deeper layer of the subject. They are designed to be read sequentially, though readers with strong backgrounds in differential geometry may enter at the second part with only occasional backward references needed.

Part I: Geometric Foundations spans four chapters and builds the mathematical language in which all subsequent physics is written. This is the foundation that makes everything else trustworthy, and the investment here pays compound interest throughout the rest of the book.

Chapter 1 introduces smooth manifolds, coordinate charts, and atlases. The key conceptual move is recognizing that a manifold is a space with a geometric identity of its own, independent of any labeling scheme one chooses to impose on it. A coordinate chart is a local labeling scheme, useful for computation but not fundamental to the geometry. Two different charts on the same manifold are two different ways of describing the same geometric reality, in the same way that two different maps of the same city are two different representations of the same streets. Tangent vectors are defined as derivations on smooth functions, a definition that makes their geometric meaning clear without any reference to coordinates. The operations that move geometric data coherently between spaces, carrying directions forward and measurements backward, are developed carefully because they are the coordinate-free origin of the transformation rules that index notation merely summarizes. Tensors are defined as multilinear maps rather than as transformation rules for arrays of numbers, a shift in viewpoint that makes their coordinate independence not a theorem to be proved but a feature built into the definition. The chapter concludes with differential forms and a geometric version of Stokes' theorem, the mechanism by which local differential laws become global conservation statements and the mathematical backbone of black hole thermodynamics.

Chapter 2 constructs connections and curvature from first principles. The central object is the Riemann curvature tensor, which measures the extent to which parallel transport around a closed loop fails to bring a vector back to itself. On a flat surface, parallel-transporting a vector around any closed loop leaves the vector unchanged. On a curved surface, the vector returns rotated by an angle proportional to the area enclosed and to the curvature of the surface. This is the geometric meaning of curvature, made precise and extended to spacetime. A careful distinction is drawn between two fundamentally different kinds of singularity: coordinate singularities, where the curvature of spacetime is completely ordinary and the problem lies entirely in the labeling scheme, and genuine curvature singularities, where the spacetime itself is torn. The singularity at the center of a Schwarzschild black hole is real in this sense: the curvature grows without bound as one approaches it, and no coordinate change can make it otherwise. The apparent singularity at the horizon in naive coordinates is not real in this sense: the curvature is perfectly finite there, and a better choice of coordinates makes the surface visibly smooth. This distinction, simple to state but easy to confuse in practice, is the single most important diagnostic tool in global black hole geometry.

Chapter 3 specializes to Lorentzian geometry, the semi-Riemannian structure required for a spacetime theory. The crucial feature of a Lorentzian metric is that it assigns different characters to different directions at each point of spacetime: timelike directions are available to massive observers moving through time, spacelike directions are available only to signals traveling faster than light and therefore to no physical object, and null directions are the paths taken by light itself. The null cone at each point of spacetime is the geometric structure that encodes which events can causally influence which others. It is, in a deep sense, the most important structure in the theory: the definition of an event horizon is a statement entirely about null cones and which of them can escape to infinity. Geodesics, which are the curves that parallel-transport their own

tangent vectors and correspond physically to the paths of freely falling observers and freely propagating light rays, are derived and analyzed. The geodesic deviation equation, which governs how neighboring freely falling trajectories accelerate toward or away from each other and thereby encodes the physical effects of tidal gravity, is one of the most useful computational tools in all of general relativity and is derived and interpreted carefully here.

Chapter 4 develops Killing fields and the geometry of symmetry. A symmetry of spacetime is not merely a qualitative statement that one direction looks the same as another. It is a precise geometric condition: a vector field is a symmetry generator if its flow preserves the metric, meaning that the spacetime looks the same after being dragged along the flow as it did before. This is the coordinate-free definition of what a physicist calls a spacetime isometry. The consequences are far-reaching: along any freely falling worldline, the inner product of the velocity vector with a symmetry generator remains constant forever. This is the geometric origin of conservation laws, not as separate physical principles appended to the theory, but as consequences of the shape of spacetime. Time-translation symmetry generates conservation of energy. Rotational symmetry generates conservation of angular momentum. A non-rotating black hole has the full symmetry of a sphere plus time translation, and these symmetries generate four independent conserved quantities along geodesics, making the problem of describing all possible orbits fully integrable. A rotating black hole loses the spherical symmetry and retains only time translation and a single axial rotation, which dramatically complicates the geodesic problem. Remarkably, a hidden additional symmetry called the Carter constant restores integrability even in the rotating case, and its discovery was one of the most important results in black hole physics in the twentieth century.

Part II: Einstein Geometry and Exact Black Hole Spacetimes moves from the mathematical language into the physical terrain. Chapter 5 presents the Einstein field equations as a geometric constraint relating the curvature of spacetime to the distribution of mass and energy within it. The equations are among the most complex in all of physics: highly nonlinear, coupled partial differential equations for the components of the metric tensor in four dimensions. Exact solutions are rare and precious. The solutions that do exist in closed form and that describe black holes are the subject of the remaining four chapters of this part.

Chapter 6 develops the Schwarzschild solution in thorough detail. The solution describes the unique spherically symmetric vacuum geometry, a result known as Birkhoff's theorem. Far from the horizon, this geometry explains the precession of Mercury's orbit, the gravitational deflection of starlight, and the gravitational redshift of signals climbing out of a gravitational well. Near and at the horizon, it introduces the phenomena that make black holes genuinely strange: the one-way membrane through which matter falls but cannot return, the relationship between the coordinates used by a freely falling observer and those used by a distant observer, and the full maximally extended spacetime revealed by a coordinate system that covers the entire geometry without degeneracy. The Penrose diagram of this maximal extension, which compresses the entire infinite spacetime into a bounded picture while preserving all causal relationships, makes the global structure transparent in a way that no other representation achieves.

Chapter 7 is devoted entirely to Penrose conformal diagrams and the systematic study of causal structure. These diagrams are among the most powerful tools in the theorist's toolkit, and they

repay careful study. The construction involves a conformal rescaling of the spacetime metric that brings all points at infinite distance to a finite location on the page, so that the causal structure of the entire infinite spacetime becomes visible at once. Penrose diagrams are constructed and interpreted for flat Minkowski spacetime, for the Schwarzschild black hole, for the Kruskal extension with its two exteriors and its white hole region, for charged black holes with their inner Cauchy horizons, and for rotating black holes. Reading a Penrose diagram fluently is not a decorative skill: it is the practical ability to reason about which observers can communicate with which others, where singularities are hidden, whether a signal sent in a given direction can ever reach infinity, and how the global causal structure changes when mass or charge is added to a system.

Chapter 8 develops the Kerr solution for rotating black holes, which are the physically relevant case since every star carries angular momentum and conservation of angular momentum means that stellar remnants, including black holes, rotate. The Kerr geometry is qualitatively richer than its non-rotating counterpart and introduces phenomena with no Schwarzschild analogue. The ergosphere is a region outside the horizon where spacetime itself rotates so rapidly that no observer, regardless of how powerful their rocket, can remain stationary relative to a distant star: they are dragged along by the rotation of the geometry itself. Within this region, the Penrose process becomes possible: an infalling particle can split into two pieces, one of which escapes to infinity carrying more energy than the original particle, and the deficit is drawn from the rotational energy of the black hole. Frame dragging, the effect by which a rotating mass pulls nearby space and gyroscopes along with it, is computed and connected to observations of orbits and spin precession around spinning compact objects. The Carter constant, the remarkable conserved quantity that restores the integrability of geodesic motion in the Kerr spacetime, is derived from a hidden symmetry structure and interpreted in terms of the separation of the wave equation on the Kerr background, a property that has made the Kerr solution the subject of more exact analytical work than almost any other spacetime.

Chapter 9 treats the charged and exotic black hole variants. The solution for a charged non-rotating black hole introduces a qualitatively new feature: a second inner horizon inside the first. Beyond this inner horizon, the mathematics suggests that an infalling observer might be able to avoid the singularity and, in the maximally extended spacetime, pass through into a different asymptotically flat region. Whether this is physically realizable or whether the inner horizon is destroyed by perturbations is the subject of the strong cosmic censorship conjecture, one of the deepest open problems in classical general relativity. The charged rotating solution, which is the most general black hole consistent with uniqueness theorems, combines all the features of the previous solutions and is analyzed in its full complexity. Black holes in spacetimes with a cosmological constant, in particular in anti-de Sitter space, are also treated here and set up the subsequent discussion of holography in Part V.

Part III: Horizon Geometry as a Field of Study is the part of the book with the least overlap with standard texts and in many respects its most original contribution. The word horizon is used freely in both physics and popular science, but its precise meaning is more subtle and more varied than almost any introductory account acknowledges.

Chapter 10 develops a careful and practically essential taxonomy. An event horizon, the concept most familiar from popular accounts, is defined as the boundary of the region of spacetime from which no light signal can ever reach a distant observer. This definition is mathematically sharp but operationally problematic: to locate the event horizon of a dynamical black hole, one needs to know the entire future history of the spacetime, including whether or not a black hole is still growing a billion years from now. Numerical relativists simulating a binary merger over a window of hundreds of milliseconds cannot locate event horizons in their simulations by this definition, because they do not have access to the spacetime's entire future. What they can locate are apparent horizons, which are defined locally as the outermost surfaces from which outgoing light rays are already beginning to converge inward rather than diverge outward. These can be found from data on a single time slice without knowing anything about the future. The relationship between these two concepts, together with the more refined notions of trapping horizons, isolated horizons, and dynamical horizons developed for different analytical purposes, is worked out carefully. The taxonomy is not pedantic: different horizon concepts satisfy different area theorems, obey different thermodynamic laws, and connect to different conserved charges. Using the wrong definition in the wrong context leads to incorrect conclusions.

Chapter 11 develops the mathematics of null congruences and derives the Raychaudhuri equation, which is arguably the single most powerful technical result in the geometry of horizons. A null congruence is a family of light rays filling a region of spacetime, the geometric object that describes the propagation of a flash of light or the expansion of a light wavefront. Three scalar quantities characterize the geometry of such a family at each point: the expansion, which measures whether the light rays are spreading apart or converging together; the shear, which measures how the cross-sectional shape of the bundle is being distorted; and the twist, which measures whether the rays are rotating around each other. The Raychaudhuri equation is a precise statement about how these quantities evolve: matter satisfying reasonable energy conditions focuses light rays, meaning it causes positive expansion to decrease and negative expansion to grow more negative. This focusing property, which reflects the gravitationally attractive nature of ordinary matter, is the geometric core of two of the most profound results in black hole physics. The first is the Penrose singularity theorem, which establishes that gravitational collapse of a sufficient mass must inevitably produce a spacetime singularity. The second is the classical black hole area theorem, which establishes that the total surface area of all black hole horizons in the universe can never decrease, provided matter behaves normally. The formal resemblance of this area theorem to the second law of thermodynamics was the clue that led Bekenstein to propose, and Hawking to confirm, that black holes carry genuine thermodynamic entropy.

Chapter 12 addresses quasi-local mass, which is one of the most technically challenging and physically important problems in the subject. In a flat spacetime, defining the energy contained within a region is straightforward: add up all the contributions. In a curved spacetime, this turns out to be impossible in a fully general and coordinate- independent way, because gravitational energy cannot be localized to a point or a bounded region. The problem is not a technical limitation awaiting a better technique; it is a fundamental consequence of the equivalence principle. Several geometric approaches to an approximate but well-defined notion of quasi-local mass are developed, each with its own range of validity and its own geometric interpretation. The Penrose inequality, one of the most important unsolved conjectures in classical general

relativity, is stated: it asserts a lower bound on the total mass of a spacetime containing a trapped surface and connects the geometry of the surface to the gravitational mass of the system in a precise and testable way.

Part IV: Geometric Methods for Dynamics and Perturbations provides the toolkit for studying black holes in situations where exact stationary solutions do not suffice. This part is essential for anyone working in numerical relativity, gravitational wave source modeling, or black hole spectroscopy.

Chapter 13 develops the decomposition of spacetime into a sequence of spatial slices evolving in time, the mathematical framework that underlies all of numerical relativity and much of the analytical study of initial value problems in general relativity. The key idea is to foliate a four-dimensional spacetime by a family of three-dimensional spatial surfaces, each representing the state of the geometry at one moment. The geometry of each slice, and the way in which the slice is embedded in the ambient spacetime, together encode all the information in the four-dimensional metric. The Einstein equations decompose into two types: constraint equations that the geometry of any given slice must satisfy, and evolution equations that determine how the geometry propagates from one slice to the next. The remaining freedom to choose how rapidly coordinate time advances and how the spatial coordinates are dragged between slices corresponds to the gauge freedom of general relativity. The development of choices that remain well-behaved during black hole formation, binary inspiral, and merger has been one of the central practical breakthroughs of numerical relativity, and this chapter provides the geometric framework for understanding why different choices behave so differently.

Chapter 14 introduces tetrads, which are locally defined orthonormal frames that allow tensor components to be measured by a specific observer rather than in a specific coordinate system. The Newman-Penrose formalism, which employs a particular type of frame adapted to the directions along which light propagates most naturally in a given spacetime, is developed as the most natural language for describing gravitational radiation. The curvature of spacetime decomposes in this frame into five complex scalars with clear physical interpretations: one of them describes outgoing gravitational radiation far from its source, directly corresponding to the signal measured by gravitational wave detectors. A theorem due to Goldberg and Sachs characterizes all the classical black hole solutions as belonging to a special class of spacetimes defined by a particular algebraic degeneracy of the Weyl curvature, a result that explains why so many exact analytical results are available for these spacetimes and not for more general ones.

Chapter 15 applies these tools to perturbation theory for black holes. When a black hole is slightly disturbed by a passing gravitational wave, an inspiraling companion, or a cloud of matter falling in, the perturbation of the spacetime geometry satisfies a linear wave equation on the unperturbed black hole background. For a non-rotating black hole, these perturbations separate cleanly into different angular modes, and each mode satisfies an equation formally identical to a quantum mechanical scattering problem in a specific potential determined by the black hole geometry. For a rotating black hole, the analogous equation is more complex but still fully separable, a non-trivial fact that is a direct consequence of the hidden symmetry structure of the Kerr spacetime discussed in Part I. The ringing frequencies at which a perturbed black hole oscillates and decays, called quasinormal modes, are the characteristic fingerprints of each black

hole's mass and spin. They appear as the ringdown signal in gravitational wave observations following binary mergers, and comparing measured ringdown frequencies to theoretical predictions tests general relativity in the strongest known gravitational fields.

Chapter 16 treats the optical geometry of black holes and the phenomena of gravitational lensing and black hole shadows. Every massive body bends light, but in the extreme gravity near a black hole, light can orbit the black hole on a circular path, forming a photon sphere. This unstable circular orbit is the geometric structure responsible for the bright emission ring visible in images of black holes produced by the Event Horizon Telescope: light that nearly orbits many times before escaping is strongly amplified, creating the photon ring, and light that falls inside the horizon creates the dark shadow at the center of the image. The relationship between the geometric properties of the photon sphere, the appearance of the shadow as seen from infinity, and the physical parameters of the black hole is derived and analyzed. A classical theorem from topology, applied to the spatial geometry seen by light rather than to the spacetime geometry directly, provides an elegant formula for the total deflection of a light ray by a massive body without requiring solution of the full geodesic equation.

Part V: Thermodynamics, Euclidean Methods, and Quantum Geometry addresses what is arguably the deepest and most mysterious convergence in all of theoretical physics: the discovery that black holes obey laws formally identical to the laws of thermodynamics, with the horizon area playing the role of entropy and the surface gravity playing the role of temperature, and the subsequent realization that this formal resemblance is not an accident but a deep identity.

Chapter 17 develops Euclidean black hole geometry, obtained by a mathematical procedure of rotating the time coordinate into imaginary values, transforming the Lorentzian spacetime metric into a purely spatial Riemannian metric. For a non-rotating black hole, this Euclidean geometry is smooth only if the imaginary time coordinate is periodic, and the required period is determined entirely by the surface gravity of the horizon. In quantum physics, periodicity in imaginary time is the precise mathematical signature of a thermal state: a system in equilibrium at a temperature inversely proportional to the period. The unavoidable periodicity of the Euclidean black hole geometry therefore implies, by purely geometric reasoning, that the black hole must be thermal. The temperature it radiates is determined entirely by its surface gravity and nothing else. This is the geometric derivation of the Hawking effect: not as a complicated calculation of particle creation by a changing gravitational field, though that calculation is also presented for completeness, but as the inevitable consequence of requiring the imaginary-time geometry to be smooth at the horizon. The result extends with full generality to rotating and charged black holes, and the corresponding temperatures are derived.

Chapter 18 treats black hole entropy. The area theorem of Chapter 11 established that the total surface area of all black hole horizons can only increase in classical general relativity. This is formally identical to the statement that total entropy can only increase in thermodynamics. The identification of horizon area with entropy, proposed by Bekenstein on information-theoretic grounds before the derivation of the Hawking temperature, is confirmed and extended. The entropy of a black hole is one quarter of its horizon area measured in units of the Planck area, a combination of the fundamental constants of quantum theory, gravity, and the speed of light. A general formula due to Wald shows that this is not a coincidence of the particular gravitational

action of general relativity, but a consequence of the symmetry of the theory under smooth coordinate transformations: any theory of gravity with this symmetry has a natural entropy formula for its black holes, and the area law is the simplest case. The microscopic origin of black hole entropy, and the question of what degrees of freedom are being counted by the area formula, is one of the deepest open problems in theoretical physics. The most developed attempts at an answer, from string theory and from loop quantum gravity, are surveyed with attention to what each has and has not succeeded in explaining.

Chapter 19 introduces anti-de Sitter spacetime and the holographic correspondence. Anti-de Sitter space is a spacetime of constant negative curvature that plays a special role in modern theoretical physics because of its geometric properties near infinity: its boundary at infinite distance has the structure of a flat spacetime of one lower dimension, and the symmetry group of the bulk anti-de Sitter space is identical to the symmetry group of a conformal field theory living on that boundary. This coincidence of symmetries is the geometric foundation of the holographic correspondence, which maps gravitational dynamics in the interior of anti-de Sitter space to non-gravitational quantum dynamics on the boundary. Black holes in anti-de Sitter space correspond to thermal states of the boundary theory, and a phase transition between hot empty space and a large black hole is the gravitational dual of a transition in the boundary theory between confined and deconfined phases. A formula relating the entanglement entropy of a region in the boundary theory to the area of a minimal surface in the interior, anchored at the boundary of that region, is derived and interpreted. This formula is one of the most important recent results in the subject because it provides a concrete geometric realization of the idea that the geometry of spacetime itself might emerge from quantum entanglement in the underlying theory.

Part VI: Extensions and Frontiers surveys several directions in which the core theory of rotating black holes is extended and tested.

Chapter 20 treats black holes in higher dimensions. Remarkably, a rotating black hole in more than four spatial dimensions is not simply a higher-dimensional version of the Kerr solution: it can rotate independently in multiple planes simultaneously, and the resulting family of solutions is qualitatively richer than anything available in four dimensions. Even more striking is the existence of extended black objects: black strings and black branes that are uniform along compact extra dimensions. These extended objects are unstable to perturbations that cause them to pinch off and fragment into sequences of roughly spherical black holes connected by thin necks, an instability discovered by Gregory and Laflamme that has no analogue in four dimensions and that illustrates how dramatically the geometry of black holes changes when extra dimensions are present.

Chapter 21 develops the Penrose and Hawking singularity theorems in their full geometric form. The Penrose singularity theorem establishes, through a remarkably general argument, that once a region of spacetime forms a trapped surface from which light rays are already converging inward in all directions, the formation of a genuine singularity is inevitable: the spacetime must become geodesically incomplete, meaning that some freely falling worldlines simply end after a finite proper time with no smooth geometric continuation. The proof uses the focusing properties of null congruences developed in Chapter 11 together with a topological argument about the structure of globally hyperbolic spacetimes. The area theorem is proved by a similar argument.

Topological censorship, which constrains the topology of spatial sections of spacetime as seen from infinity, is established and its implications for the possible shapes of black holes in higher dimensions are discussed.

Chapter 22 addresses modified gravity theories. General relativity is the unique metric theory of gravity consistent with a set of elegant geometric assumptions, but it is widely expected to require modification at very short distances where quantum effects become important. Corrections to the gravitational action involving higher powers of curvature arise naturally in attempts to quantize gravity, and these corrections modify the geometry of black holes in specific and computable ways. Scalar-tensor theories, in which a new scalar field couples non-minimally to the curvature of spacetime, allow for spontaneous scalarization of black holes, a phenomenon in which the standard solution without a scalar field becomes unstable and the black hole develops scalar hair under certain conditions. Theories with torsion and non-metricity, in which the rules connecting the metric to the connection are relaxed, provide geometric frameworks for exploring the space of possible gravitational theories. For each class of modifications, the chapter develops the geometric framework, identifies how the black hole solutions change, and discusses what observations could in principle distinguish the modified theory from standard general relativity.

Chapter 23 presents numerical relativity as an applied branch of computational differential geometry. The remarkable success of numerical relativity in producing accurate gravitational wave predictions from binary black hole simulations, predictions that have been matched against observations with extraordinary precision, rests entirely on a set of geometric choices: how to slice spacetime into spatial surfaces, how to choose coordinates that remain regular near the horizon throughout a long evolution, how to locate and track the apparent horizon as a geometric surface, and how to extract the outgoing gravitational wave signal in a coordinate-independent way at large distances. This chapter traces each of these choices back to the geometric foundations developed in earlier parts of the book. The apparent horizon finder is a geometric optimization problem on a curved surface. The gravitational wave extraction formula is a statement about the asymptotic falloff of specific curvature components in a particular frame. The gauge choices that have made long binary merger simulations stable since 2005 have geometric interpretations in terms of which slicing conditions keep the coordinate system from being pulled into the singularity. The chapter closes the circle opened in Chapter 1: the abstract ideas of manifolds, coordinate-free tensors, and geometric invariants are here the operational tools of one of the most computationally demanding and observationally productive fields in modern physics.

Data-Driven Interludes: Connecting the Abstract to the Observable

A feature of this text that distinguishes it from purely mathematical treatments is the inclusion, at strategic points in the development, of data-driven interludes that ground the geometric abstractions in real observational data.

The first and most extended of these interludes occupies a substantial section of Chapter 1. It analyzes public gravitational wave event catalogs, drawing on the three major data releases from the gravitational wave observatory network to assemble a sample of 88 high-confidence compact binary merger events in which the primary object has a mass well above the maximum possible for a neutron star. From the catalog's reported parameter estimates, the analysis computes and displays a series of scalar quantities: the masses of the two merging bodies as measured in the frame of the source rather than the frame of the detector, the chirp mass which is the specific combination of the two masses that most directly controls how rapidly the orbit decays and how loudly the system radiates, the effective inspiral spin parameter which measures the component of the spin angular momenta aligned with the orbital angular momentum, the mass ratio, the fraction of the total mass radiated as gravitational waves during the merger, and the final mass of the remnant black hole.

Each of these quantities is coordinate-free in a precise and important sense: it is a scalar invariant of the binary system within a specified physical model, not a component of a tensor that depends on a choice of frame. The figures serve a dual purpose. At the level of astrophysics, they reveal genuine population-level structure in the observed black hole mergers: a clustering of events in the several-tens-of- solar-masses range, a predominance of nearly equal-mass binaries, a selection effect that makes more massive systems detectable at larger distances, and a remarkably consistent radiated mass fraction of a few percent of the total mass across the entire observed population. At the conceptual level, they make a simple point that is easy to miss but important not to: the coordinate-free scalar invariants being introduced in the theory sections of the chapter are precisely the quantities that appear in the data products released by the observational collaborations. The abstract is not separate from the practical; it is what the practical is built on.

Subsequent data interludes appear in later chapters to connect quasi-local mass to observations, black hole shadow geometry to Event Horizon Telescope images, and quasinormal mode frequencies to the ringdown signals in gravitational wave data.

Prerequisites and Intended Readers

This book is aimed at advanced graduate students in theoretical physics who have completed at least one graduate course in general relativity, and at researchers in adjacent fields who wish to develop a deeper geometric understanding of black hole physics than standard introductory treatments provide.

The assumed prior knowledge includes familiarity with the basic exact solutions of general relativity, comfort with the rules of tensor index manipulation and the summation convention, and some exposure to the Einstein field equations and their derivation from a variational principle. What is not assumed, and what Part I builds systematically from the beginning, is fluency in coordinate-free differential geometry: the language of smooth manifolds, tangent and cotangent bundles, exterior algebra, Lie derivatives, and connections as abstract geometric structures rather than as transformation rules for components.

Readers with strong mathematical backgrounds in differential or Riemannian geometry approaching the subject from the mathematics side will find the physical motivation and the data interludes provide the astrophysical context that mathematical monographs typically leave implicit. Readers from numerical relativity or gravitational wave astronomy who bring strong computational skills but less familiarity with geometric formalism will find that Parts I through III supply the theoretical underpinning for tools and choices they may already use in practice without a complete geometric justification.

Some familiarity with quantum field theory is helpful for Part V but is not required in advance. The arguments are kept as geometric as possible, and the field-theoretic concepts that are needed, including the relationship between imaginary-time periodicity and thermal states, are introduced in context rather than assumed.

A Note on Mathematical Style

Every definition in this book is stated with precision. But every definition is also accompanied by an explanation of what geometric phenomenon it is designed to capture, why it takes the specific form it does, and what would be misidentified or missed if it were weakened or replaced by a simpler substitute. Every theorem is accompanied by an account of its geometric content and of why its hypotheses are not merely technical conditions but genuine reflections of properties of the spacetimes to which it applies.

Proofs are given in full where the argument itself illuminates the geometry. They are given in outline where the key insight can be stated clearly and the remaining technical steps, while necessary, are standard and do not add conceptual content. References to the research literature are provided in either case with enough specificity to serve as a genuine guide to further reading rather than a gesture toward academic convention.

A Final Word

The geometry of black holes repays the effort of learning it properly, and that effort is real. The formalism is demanding, the definitions are numerous and carefully drawn, and the temptation to abbreviate the geometric development and retreat to a more computational style is always present. The book has resisted that temptation deliberately, because the reward for doing so is access to a way of thinking about spacetime that is not merely mathematically reliable but genuinely illuminating.

When the coordinate-free picture is fully internalized, things that seemed mysterious become clear. The event horizon stops being a place where a number in a formula becomes infinite and becomes instead a comprehensible causal surface: the outermost boundary beyond which the future of every event is confined to the black hole interior. The singularity theorems stop being difficult analytical results and become almost obvious geometric statements about what happens

when light rays are forced to converge. The area theorem becomes a natural consequence of the fact that matter focuses light, applied to the null generators of the event horizon. The Hawking temperature becomes the inevitable outcome of requiring a certain geometric smoothness. The entropy formula becomes the Noether charge of a symmetry. Each of these results, seen from the right vantage point, has the quality of something that could not have been otherwise.

There is a specific and deep pleasure in reaching that vantage point on a difficult subject. This book is an attempt to build the path toward it carefully, from foundations that can support the full weight of what is built on them, and to maintain throughout the conviction that real understanding and mathematical rigor are not in tension with each other but are, at their best, the same achievement approached from different angles.

Black Holes Geometry: Curvature and Horizons

Contents

Book Roadmap Index Table **iii**

I Geometric Foundations **1**

1 Manifolds, Tensors, and Physics as Coordinate Free Thought **2**

1.1 Introduction . . . 3

1.2 Smooth manifolds, charts, and atlases: why coordinates are a polite lie . . . 3

1.2.1 A first conceptual visual: an atlas as a cover by local stories . . . 3

1.3 Tangent vectors, pushforwards, and direction without coordinates . . . 4

1.4 Cotangent vectors, pullbacks, and how covariant objects travel . . . 5

1.5 Tensors as multilinear maps . . . 5

1.5.1 Coordinate free viewpoint: tensors as maps, components as representations 6

1.6 Vector fields, flows, diffeomorphisms, and Lie brackets . . . 6

1.7 Differential forms, wedge product, and exterior derivative . . . 7

1.8 Stokes theorem as the globalization engine . . . 8

1.9 Data driven interlude: coordinate free scalars from gravitational wave black hole catalogs . . . 8

1.9.1 Dataset provenance and minimal processing . . . 9

1.9.2 Illustration 1: component masses in the source frame . . . 10

1.9.3 Illustration 2: chirp mass distribution . . . 10

1.9.4 Illustration 3: effective inspiral spin distribution . . . 10

1.9.5 Illustration 4: luminosity distance versus chirp mass . . . 10

1.9.6 Illustration 5: cumulative detections versus time . . . 12

1.9.7 Illustration 6: mass ratio distribution . . . 12

1.9.8 Illustration 7: radiated mass fraction distribution . . . 12

1.9.9 Illustration 8: final mass versus total mass 12
1.9.10 Illustration 9: redshift versus luminosity distance 15
1.9.11 Illustration 10: network matched filter signal to noise ratio 15
1.10 Real world phenomena: why coordinate free tools matter 15
1.11 Conclusion 17

Book Roadmap Index Table

Part	Chapter	Title (planned)
Part I: Geometric Foundations	1	Manifolds, Tensors, and Physics as Coordinate Free Thought
Part I: Geometric Foundations	2	Connections, Curvature, and Parallel Transport
Part I: Geometric Foundations	3	Lorentzian Geometry: Time Joins the Chat
Part I: Geometric Foundations	4	Symmetries as Geometry: Killing Fields and Conserved Quantities
Part II: Einstein Geometry and Exact Black Hole Spacetimes	5	Einstein's Equations as Curvature Constraints
Part II: Einstein Geometry and Exact Black Hole Spacetimes	6	Schwarzschild Geometry: The Simplest Nontrivial Monster
Part II: Einstein Geometry and Exact Black Hole Spacetimes	7	Global Structure and Penrose Diagrams: Causality as Artwork
Part II: Einstein Geometry and Exact Black Hole Spacetimes	8	Kerr Geometry: Rotation Warps Reality in a New Dimension
Part II: Einstein Geometry and Exact Black Hole Spacetimes	9	Charged and Exotic Variants: Kerr-Newman and Friends
Part III: Horizon Geometry as a Field of Study	10	Horizons: Event, Apparent, Trapping, Isolated, Dynamical

Part	Chapter	Title (planned)
Part III: Horizon Geometry as a Field of Study	11	Null Congruences and Raychaudhuri: Focusing as Theorem Machine
Part III: Horizon Geometry as a Field of Study	12	Quasi-local Mass and Geometric Inequalities
Part IV: Geometric Methods for Dynamics and Perturbations	13	Three plus One Geometry and the Initial Value Problem
Part IV: Geometric Methods for Dynamics and Perturbations	14	Tetrads, Spinors, and Newman-Penrose Geometry
Part IV: Geometric Methods for Dynamics and Perturbations	15	Perturbations of Black Hole Geometry
Part IV: Geometric Methods for Dynamics and Perturbations	16	Optical Geometry, Lensing, and Finsler or Randers Tricks
Part V: Thermodynamics, Euclidean Methods, and Quantum Geometry	17	Euclidean Black Hole Geometry and Temperature
Part V: Thermodynamics, Euclidean Methods, and Quantum Geometry	18	Entropy, Area, and Geometric Microstories
Part V: Thermodynamics, Euclidean Methods, and Quantum Geometry	19	Holography and Hyperbolic Geometry: AdS as a Geometry Playground
Part VI: Extensions and Frontiers	20	Higher Dimensional Black Hole Geometries
Part VI: Extensions and Frontiers	21	Topology, Censorship, and Singularity Theorems
Part VI: Extensions and Frontiers	22	Modified Gravity and Non-Riemannian Geometry: Carefully Labeled
Part VI: Extensions and Frontiers	23	Numerical Relativity as Computational Differential Geometry

Part I

Geometric Foundations

Chapter 1

Manifolds, Tensors, and Physics as Coordinate Free Thought

Abstract

Black hole spacetimes are often introduced through metrics written in specific coordinates, but the mathematics of curvature and horizons does not live in those coordinates. It lives on manifolds, in tensor fields, and in differential forms: objects that keep their meaning when the coordinate labels change, disappear, or become singular. This chapter builds a coordinate free toolkit for the rest of the book. We introduce smooth manifolds, charts, and atlases and explain why coordinates can be a polite lie: they are useful interfaces, not the underlying geometry. We then develop tensors as multilinear maps, along with the pushforward and pullback operations that let geometric data move between manifolds. Vector fields, flows, diffeomorphisms, and Lie brackets are presented as the language of change that does not depend on a particular grid. Differential forms and the exterior derivative are introduced as the natural calculus for conservation laws, fluxes, and circulation, culminating in Stokes' theorem, which turns local differential information into global integral statements.

To anchor the abstractions in real measurements, a data driven interlude analyzes open gravitational wave catalogs of compact binary mergers that include many black hole systems. The figures in that interlude are produced from released observational parameter estimates and simple derived quantities such as mass ratio, chirp mass consistency checks, and estimated radiated mass fractions. The chapter ends with application focused phenomenon statements showing how coordinate free thinking becomes a practical engineering and astrophysical tool.

Keywords: smooth manifolds, charts and atlases, tensors, pullback and pushforward, vector fields, Lie bracket, differential forms, exterior derivative, Stokes theorem, gravitational wave catalogs

1.1 Introduction

General relativity can be read in two ways. The first way is as a list of coordinate expressions: write a metric $g_{\mu\nu}(x)$, compute Christoffel symbols $\Gamma^{\rho}{}_{\mu\nu}(x)$, and then chase indices until a curvature invariant pops out. The second way is as a story about geometric objects that exist without coordinates: a Lorentzian manifold $(\mathcal{M}, g)$, a curvature tensor R, and hypersurfaces whose causal character tells you whether they can be crossed by light. The first way is efficient in the short term, but it becomes fragile when coordinates misbehave. Horizons, null surfaces, and global causal structure are precisely where naive coordinates tend to lie to you, for example by hiding regular regions behind coordinate singularities. The second way is slower at the beginning, but once learned it makes later chapters stable and conceptually clean [1, 4, 5].

Coordinate free thought is not a philosophical luxury. It is a practical form of error control. In data analysis, coordinate dependence is a common failure mode: a quantity that appears to vary can be an artifact of how you labeled points. In black hole physics, the same idea appears as gauge dependence. Many intermediate calculations depend on choices that are not physical, such as a coordinate chart, a tetrad, or a slicing. The final predictions that can be compared to telescopes and detectors must not depend on those choices. The aim of this chapter is to introduce the basic geometric operators that enforce that independence.

1.2 Smooth manifolds, charts, and atlases: why coordinates are a polite lie

A smooth manifold is a space that looks locally like $\mathbb{R}^n$ but may have global structure that cannot be captured by a single coordinate patch. This statement is the first place where a black hole intuition already helps. The event horizon of a stationary black hole is not a point or a curve; it is a null hypersurface, a three dimensional submanifold of the full spacetime. Locally it behaves like $\mathbb{R}^3$, but globally it can have nontrivial topology and causal behavior, and the coordinate systems that make computations easy in one region often become awkward elsewhere.

Definition 1.1 (Smooth manifold)**.** A *smooth n dimensional manifold* $\mathcal{M}$ is a set equipped with an atlas of charts $\{(U_\alpha, \varphi_\alpha)\}$ such that each $U_\alpha \subset \mathcal{M}$ is open, $\varphi_\alpha : U_\alpha \to \mathbb{R}^n$ is a homeomorphism onto its image, and on overlaps $U_\alpha \cap U_\beta$ the transition map $\varphi_\beta \circ \varphi_\alpha^{-1}$ is smooth wherever it is defined [1].

The power of the definition is not the word smooth; it is the word *overlaps*. A single chart is just a labeling scheme. An atlas forces compatibility between different labeling schemes. The geometry lives in the compatibility, not in any one set of coordinates. This is why coordinates can be a polite lie. They politely pretend that the space is $\mathbb{R}^n$ everywhere, while the atlas keeps track of when that pretense fails and how to translate between local pretenses.

1.2.1 A first conceptual visual: an atlas as a cover by local stories

The same manifold can be described by many atlases. In physics we almost always work with atlases that respect symmetries, such as spherical symmetry for Schwarzschild or stationarity

and axial symmetry for Kerr. The underlying manifold does not care. The atlas is a choice of language.

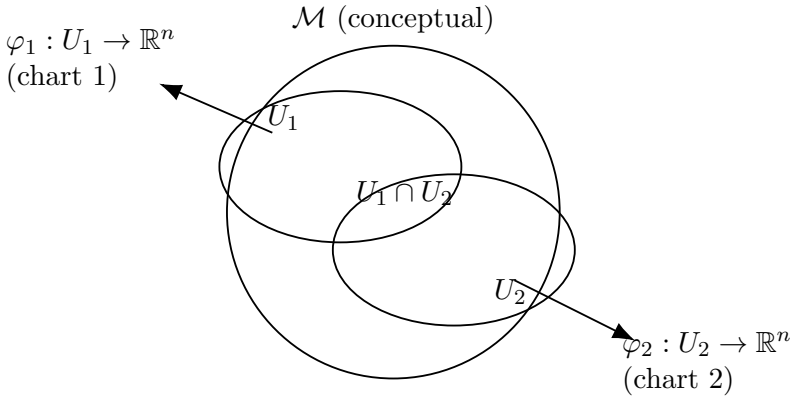


Figure 1.1: Conceptual atlas view: two overlapping charts covering the same manifold. A chart is a local coordinate story. The overlaps enforce compatibility, which is where geometry becomes coordinate independent [1].

In later chapters, this picture reappears in more serious form. Kruskal coordinates for Schwarzschild and Boyer-Lindquist coordinates for Kerr are not alternative physical universes. They are different charts on (possibly different extensions of) the same geometric spacetime. The moment you understand that, the statement "$r = 2M$ is not a singularity" becomes natural rather than mysterious [4, 5].

1.3 Tangent vectors, pushforwards, and direction without coordinates

A manifold is not a vector space. You cannot add points. Yet physics needs the idea of direction, velocity, and infinitesimal change at a point. The tangent space $T_p\mathcal{M}$ is the device that restores linear structure locally. There are multiple equivalent definitions: as equivalence classes of curves, as directional derivatives acting on smooth functions, or as derivations. The derivation viewpoint is the most coordinate free and is the one that scales best to later constructions [1].

Definition 1.2 (Tangent vector as derivation)**.** Let $p \in \mathcal{M}$. A tangent vector $v \in T_p\mathcal{M}$ can be defined as a linear map $v : C^\infty(\mathcal{M}) \to \mathbb{R}$ such that for all $f, g \in C^\infty(\mathcal{M})$,

$$v(fg) = f(p)\, v(g) + g(p)\, v(f).$$

This Leibniz rule is what makes v behave like a directional derivative at p without requiring any coordinates. If you later choose a chart (U, φ), you can express v in components as $v = v^\mu \partial_\mu$, but that component list is a representation, not the object.

Now consider a smooth map $F : \mathcal{M} \to \mathcal{N}$ between manifolds. How does a direction at $p \in \mathcal{M}$ map to a direction at $F(p) \in \mathcal{N}$? The answer is the pushforward, sometimes called the differential of F.

Definition 1.3 (Pushforward)**.** Given $F : \mathcal{M} \to \mathcal{N}$ smooth and $p \in \mathcal{M}$, the pushforward

$F_* : T_p\mathcal{M} \to T_{F(p)}\mathcal{N}$ is defined by

$$(F_* v)(h) = v(h \circ F), \qquad h \in C^\infty(\mathcal{N}).$$

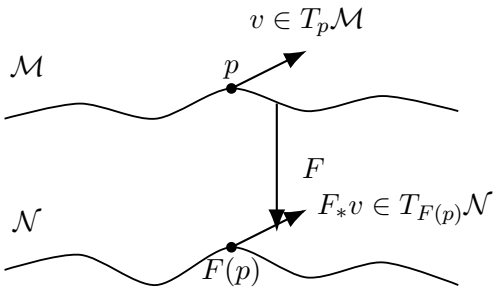


Figure 1.2: Conceptual pushforward: the map F transports a direction v at p into a direction $F_* v$ at the image point. In coordinate calculations this becomes a Jacobian acting on components, but the definition does not depend on coordinates [1].

This operation is the backbone of coordinate transformations and of symmetry actions. When we later treat isometries and Killing fields, we will use pushforwards to express how a diffeomorphism moves geometric data while preserving the metric [5].

1.4 Cotangent vectors, pullbacks, and how covariant objects travel

If tangent vectors are the language of directions, then cotangent vectors are the language of linear measurements of those directions. The cotangent space $T_p^*\mathcal{M}$ is the dual vector space of $T_p\mathcal{M}$. A covector $\alpha \in T_p^*\mathcal{M}$ eats a tangent vector and returns a number. In physics that number can be a differential change in a scalar field, a work term, or a piece of momentum.

Maps between manifolds naturally pull covectors backward rather than push them forward. This is the origin of why indices behave the way they do. Contravariant objects push forward, covariant objects pull back.

Definition 1.4 (Pullback)**.** Given $F : \mathcal{M} \to \mathcal{N}$ smooth, the pullback of a covector $\beta \in T_{F(p)}^*\mathcal{N}$ is the covector $F^*\beta \in T_p^*\mathcal{M}$ defined by

$$(F^*\beta)(v) = \beta(F_* v), \qquad v \in T_p\mathcal{M}.$$

This pullback is what makes differential forms especially powerful in physics. Maxwell's equations, conservation laws, and flux integrals are naturally expressed in a way that survives any coordinate system used to compute them [3, 6].

1.5 Tensors as multilinear maps

A tensor is a device that takes several vectors and covectors as inputs and produces a number, in a way that is linear in each slot. The modern viewpoint is not that a tensor is an array of

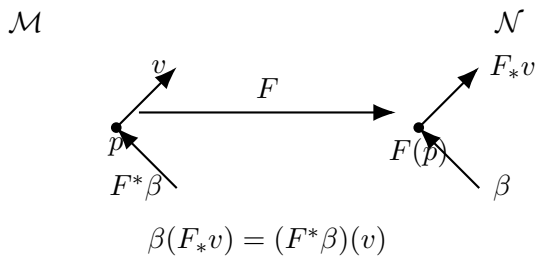


Figure 1.3: Conceptual pullback: covectors and, more generally, differential forms are transported by composing with the pushforward. This is the coordinate free origin of the transformation rules for covariant indices [1, 3].

numbers with many indices, but that it is a multilinear map whose component arrays appear only after choosing bases.

Definition 1.5 (Tensor of type (r, s))**.** Let $p \in \mathcal{M}$. A tensor of type (r, s) at p is a multilinear map

$$T : \underbrace{T_p^*\mathcal{M} \times \cdots \times T_p^*\mathcal{M}}_{r \text{ times}} \times \underbrace{T_p\mathcal{M} \times \cdots \times T_p\mathcal{M}}_{s \text{ times}} \to \mathbb{R}$$

that is linear in each argument [1].

The metric g is a tensor of type $(0, 2)$: it eats two vectors and returns their inner product. The curvature tensor R is type $(1, 3)$ or $(0, 4)$ depending on conventions. The stress energy tensor T_{ab} is also type $(0, 2)$. When you write these objects with indices, you are describing their action in a particular basis. The tensor itself does not change under coordinate transformations, even though its component array does.

1.5.1 Coordinate free viewpoint: tensors as maps, components as representations

The most reliable way to remember how tensors transform is to not memorize component rules at all. Instead, remember the definition as a multilinear map. If you change coordinates, you change the basis vectors and covectors used to represent that map, so the component array changes, but the underlying mapping does not. In later chapters we will repeatedly exploit this: we compute in whatever chart is convenient, then confirm that the final statement is about a tensorial object or a scalar invariant [1, 4].

1.6 Vector fields, flows, diffeomorphisms, and Lie brackets

A single tangent vector lives at a single point. A vector field assigns a tangent vector to every point in a region. Formally, a smooth vector field is a smooth section of the tangent bundle, $X \in \Gamma(T\mathcal{M})$. Vector fields are the geometry of infinitesimal motion. They generate flows, which are one parameter families of diffeomorphisms. Diffeomorphisms are smooth invertible maps with smooth inverses, and in general relativity they represent the gauge freedom of coordinate choice.

Definition 1.6 (Flow of a vector field)**.** Let X be a smooth vector field. A (local) flow is a map $\Phi : \mathbb{R} \times \mathcal{M} \to \mathcal{M}$ such that $\Phi(0, p) = p$ and $\frac{d}{dt}\big|_{t=0}\Phi(t, p) = X_p$, whenever the derivative makes sense [1].

The Lie bracket measures the failure of flows to commute. If you flow along X and then along Y, and compare to the reverse order, the first nontrivial difference is captured by $[X, Y]$.

Definition 1.7 (Lie bracket)**.** For vector fields X, Y, their Lie bracket $[X, Y]$ is the vector field defined by its action on functions:

$$[X, Y](f) = X(Y(f)) - Y(X(f)).$$

This bracket is not merely algebra. In black hole spacetimes, the commutation properties of symmetry generators determine conserved quantities, integrability, and whether you can adapt coordinates to a family of observers. Later, Killing fields are defined by the Lie derivative of the metric vanishing, a statement expressed entirely in terms of flows and pushforwards [4, 5].

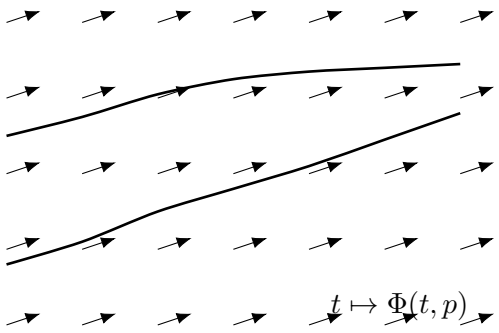


Figure 1.4: Conceptual vector field and flow lines: a smooth vector field assigns a direction at each point, and its integral curves describe the flow it generates. In spacetime geometry, such flows describe families of observers or symmetry transformations [1, 5].

1.7 Differential forms, wedge product, and exterior derivative

Differential forms generalize the differential *df* of a scalar field, line integrals, surface integrals, and volume integrals into a unified coordinate free calculus. A k form assigns a number to k tangent vectors and is multilinear and alternating. The alternating property builds orientation into the object. This is why forms are the natural language for flux and circulation.

Definition 1.8 (Differential k form)**.** A smooth differential k form ω on $\mathcal{M}$ assigns to each $p \in \mathcal{M}$ an alternating multilinear map

$$\omega_p : \underbrace{T_p\mathcal{M} \times \cdots \times T_p\mathcal{M}}_{k \text{ times}} \to \mathbb{R},$$

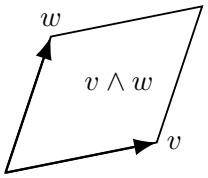
varying smoothly with p [2, 3].

The wedge product $\wedge$ is an antisymmetric multiplication that turns a k form and an ℓ form into a $(k + \ell)$ form. The exterior derivative d increases degree by one and is defined by two key

requirements: it agrees with the usual differential on functions, and it satisfies $d^2 = 0$ along with a graded Leibniz rule. These axioms fix it uniquely.

Proposition 1.1 (Key identities for exterior calculus)**.** Let α be a k form and β an ℓ form. Then

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta, \qquad d(d\alpha) = 0.$$



A 2 form ω assigns a signed area weight $\omega(v, w)$

Figure 1.5: Conceptual wedge intuition: an oriented 2 form evaluates on an ordered pair of tangent vectors and produces a signed area density. Orientation and antisymmetry are built in, which is why forms naturally encode flux and circulation [2, 3].

In black hole astrophysics, the magnetized plasma of an accretion flow is governed by equations where magnetic flux conservation can be written elegantly with differential forms. Even in flat space, writing electromagnetism with forms clarifies what is coordinate dependent and what is intrinsic. In curved spacetime this clarity becomes essential [3, 6].

1.8 Stokes theorem as the globalization engine

If the exterior derivative is the local operator, Stokes theorem is the global bridge. It states that integrating $d\omega$ over a region equals integrating ω over the boundary. In one sentence it unifies the fundamental theorem of calculus, Green's theorem, the divergence theorem, and the classical Stokes theorem of vector calculus.

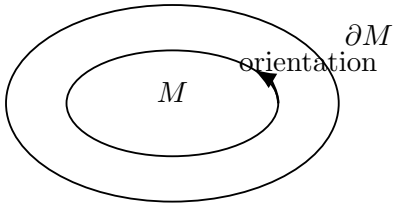
Proposition 1.2 (Stokes theorem)**.** Let M be an oriented smooth manifold with boundary ∂M, and let ω be a compactly supported $(k-1)$ form on M with $k = \dim M$. Then

$$\int_M d\omega = \int_{\partial M} \omega.$$

The phrase globalization engine is not hype. Local differential equations like $dF = 0$ or $\nabla_a T^{ab} = 0$ become global conservation laws only when you integrate and invoke Stokes. The geometry of the boundary, including its orientation and causal character in spacetime, determines what you can say globally. Horizons, being null hypersurfaces, demand extra care precisely because the induced metric degenerates. That subtlety will appear later, but the mathematical backbone is already here [4, 5].

1.9 Data driven interlude: coordinate free scalars from gravitational wave black hole catalogs

The rest of this chapter is abstract by design, but abstraction is not a replacement for reality. It is a control system that helps us extract reality from noisy measurements. To make that concrete,



$$\int_M d\omega = \int_{\partial M} \omega$$

Figure 1.6: Conceptual Stokes theorem picture: local information ($d\omega$) integrated over a region equals boundary information (ω) integrated over the boundary. This is the mechanism by which local differential laws become global integral constraints [2, 3].

we analyze open catalogs of gravitational wave events from compact binary coalescences. Many of these events are binary black hole mergers, which means the data are already entangled with the strongest curvature regimes we will model later. The gravitational wave strain is a tensorial observable, but the public catalogs distill the observations into scalar parameters such as source frame masses and effective inspiral spin. These scalars are not coordinate components; they are coordinate free summaries of the physical system within a modeling framework [7, 11–13].

1.9.1 Dataset provenance and minimal processing

We use the Gravitational Wave Open Science Center event tables for the GWTC-1-confident, GWTC-2.1-confident, and GWTC-3-confident releases [8–10]. For each event we take the median values of key inferred parameters. When a common event appears in multiple releases, we keep the latest release entry. This yields the canonical set of 90 high probability events summarized in GWTC-3 [13]. To focus on black hole containing systems, we restrict attention to events with source frame primary mass $m_1 \geq 3\,M_\odot$, which keeps 88 events. We then compute three derived scalar quantities:

$$q = \frac{\min(m_1, m_2)}{\max(m_1, m_2)}, \qquad \eta = \frac{m_1 m_2}{(m_1 + m_2)^2}, \qquad f_{\rm rad} = 1 - \frac{M_f}{M_{\rm tot}},$$

where M_f is the inferred final mass and $M_{\rm tot}$ is the inferred total mass in the source frame. These derived values are coordinate free functions of the inferred scalars.

Data note: what is coordinate free and what is model dependent

The quantities plotted below are coordinate independent in the sense that they are not tied to a particular coordinate chart on spacetime. However, they are not free of assumptions. Converting detector strain to source frame masses involves waveform modeling and cosmological inference. The point of this interlude is not to hide those assumptions; it is to show how, once a model is fixed, geometric thinking emphasizes invariants rather than coordinate artifacts [12, 13].

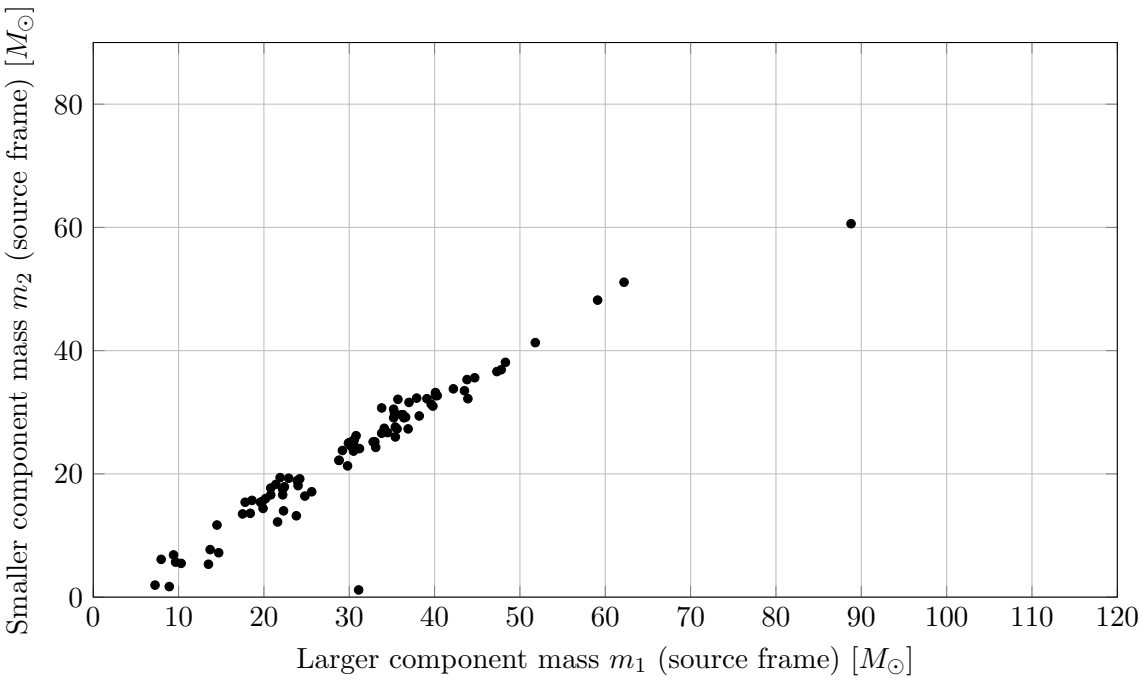


Figure 1.7: Source frame component masses for 88 black hole containing mergers from the GWTC-3 cumulative confident set. Each point is a catalog median. The plot is a coordinate free summary because masses are scalars, not coordinate components [8–10, 13].

1.9.2 Illustration 1: component masses in the source frame

The population is not uniform. There is a dense band of events with m_1 in the few tens of solar masses and m_2 somewhat smaller, and there are notable outliers at higher masses. A key geometric lesson already appears: the mass plane is not a coordinate plane on spacetime. It is a parameter space, a different kind of manifold that exists because we compress complicated tensor data into a handful of invariants. Later, when we discuss moduli spaces of solutions and parameter estimation, this viewpoint becomes central.

1.9.3 Illustration 2: chirp mass distribution

1.9.4 Illustration 3: effective inspiral spin distribution

1.9.5 Illustration 4: luminosity distance versus chirp mass

A simple statistic makes the selection effect quantitative. The correlation between $\mathcal{M}$ and $\log_{10} D_L$ in this subset is positive (about 0.67 in a Pearson correlation computed on the released medians). The exact value is not universal, but the existence of a strong correlation is a reminder that observed distributions are rarely intrinsic distributions.

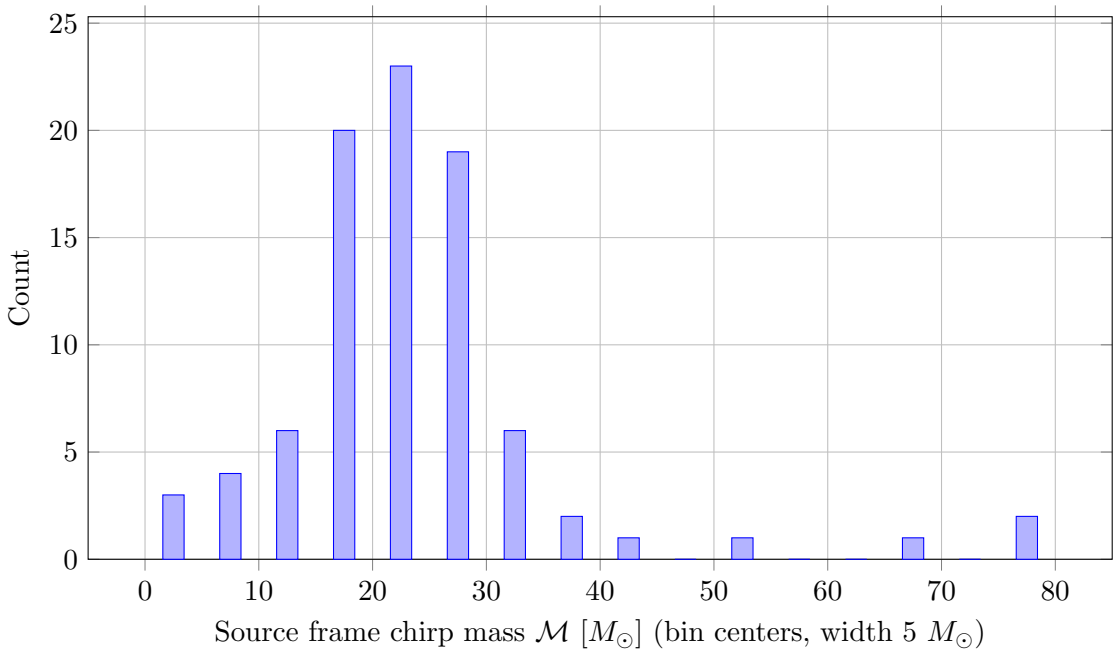


Figure 1.8: Histogram of source frame chirp mass for the same 88 events. The chirp mass is a scalar combination of the two masses that directly controls the leading inspiral phase evolution, illustrating how dynamics often depends on invariants rather than on coordinate choices [12, 13].

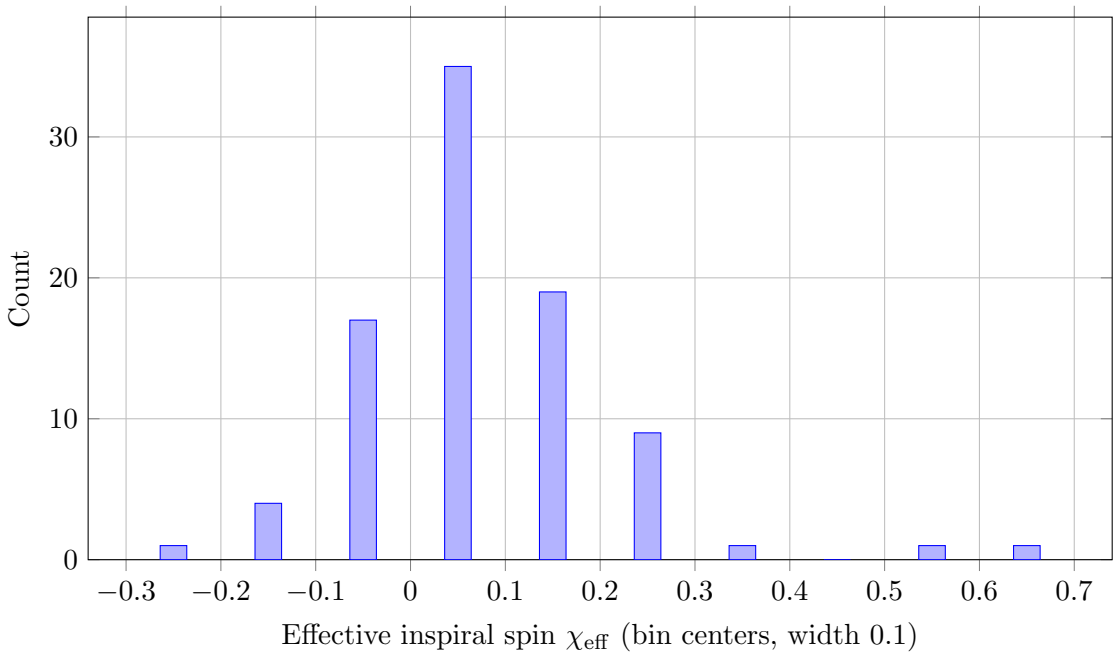


Figure 1.9: Histogram of χ_{eff} for the 88 events. The distribution is strongly peaked near small positive values, with a tail to negative values. The key point here is geometric: χ_{eff} is designed to be a coordinate free combination of spin projections along the orbital angular momentum [12, 13].

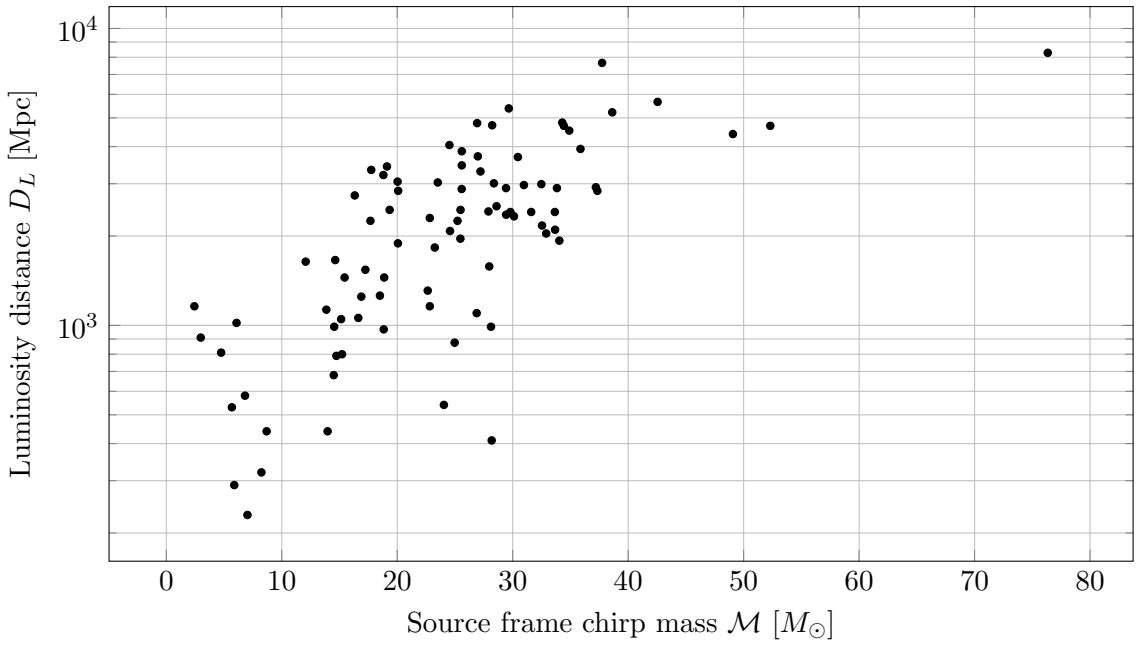


Figure 1.10: Distance versus chirp mass for 88 events (log scale in distance). The visible trend is a selection effect: more massive binaries can be detected farther away. Interpreting such trends correctly requires separating coordinate artifacts from true invariants, a recurring theme in geometric physics [12, 13].

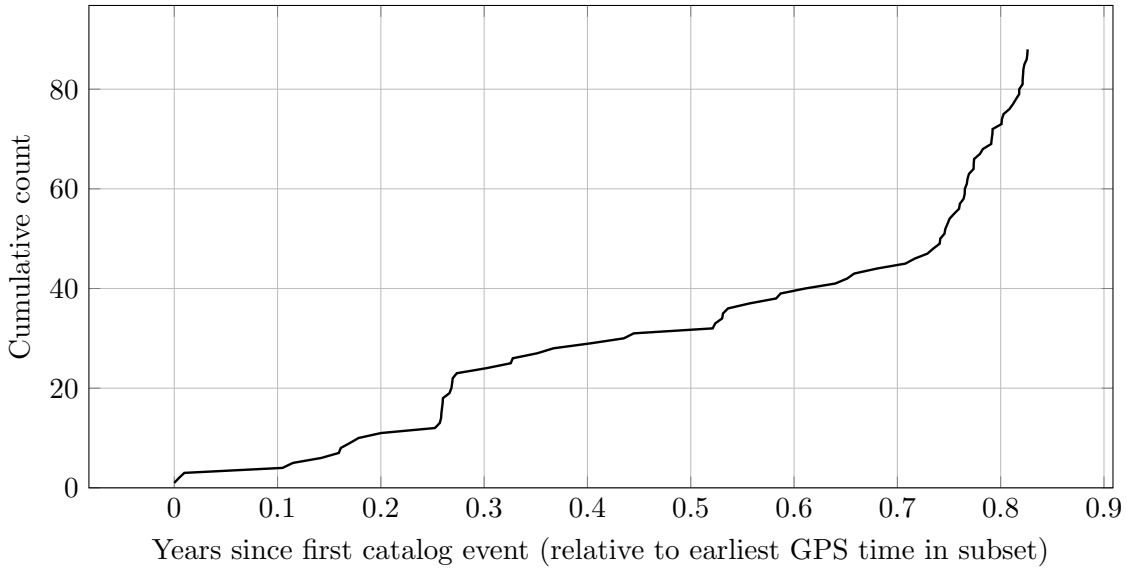


Figure 1.11: Cumulative number of black hole containing events as a function of relative time (years) computed from GPS timestamps. Using relative time is itself a coordinate change on the one dimensional time manifold: it removes a large offset without changing the geometric content of ordering [8–10].

1.9.6 Illustration 5: cumulative detections versus time

1.9.7 Illustration 6: mass ratio distribution

1.9.8 Illustration 7: radiated mass fraction distribution

1.9.9 Illustration 8: final mass versus total mass

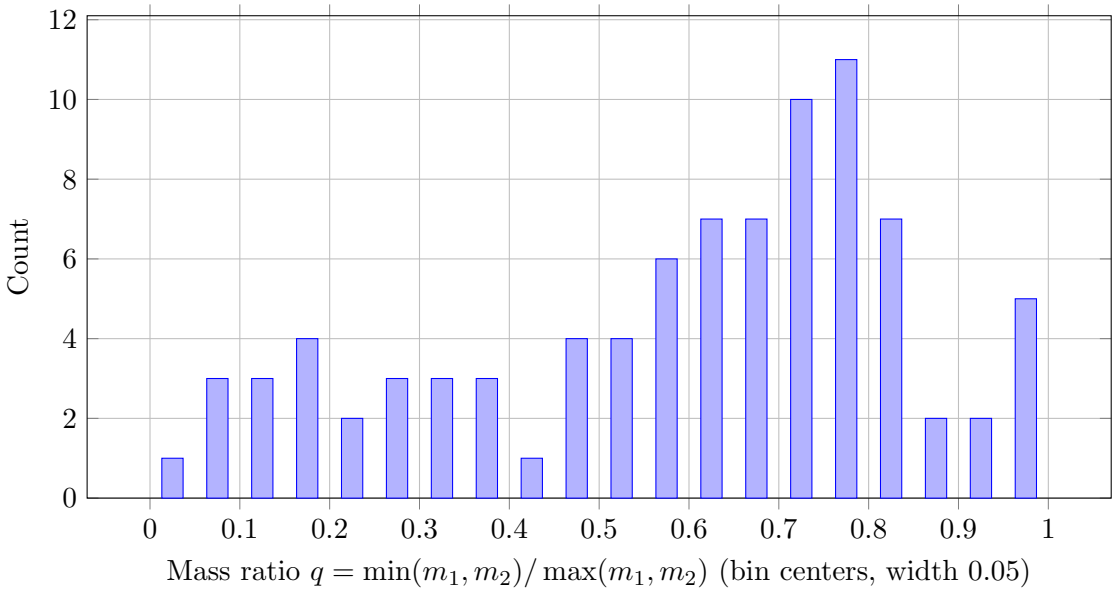


Figure 1.12: Histogram of the mass ratio q computed from source frame component masses. This is a derived scalar on the parameter manifold. Extreme mass ratio events are rare but present, and they motivate later discussions of perturbative methods and tidal geometry [13].

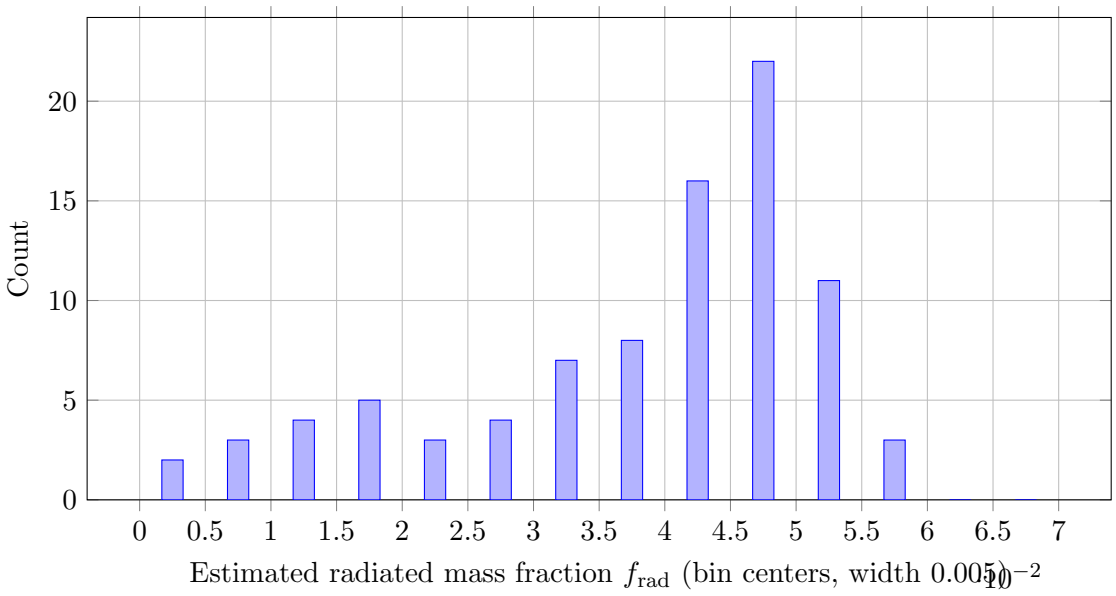


Figure 1.13: Histogram of $f_{\mathrm{rad}} = 1 - M_f/M_{\mathrm{tot}}$ using catalog medians for final and total mass (source frame). The distribution clusters around a few percent, consistent with the idea that mergers convert a small but astrophysically significant fraction of mass into gravitational radiation [12, 13].

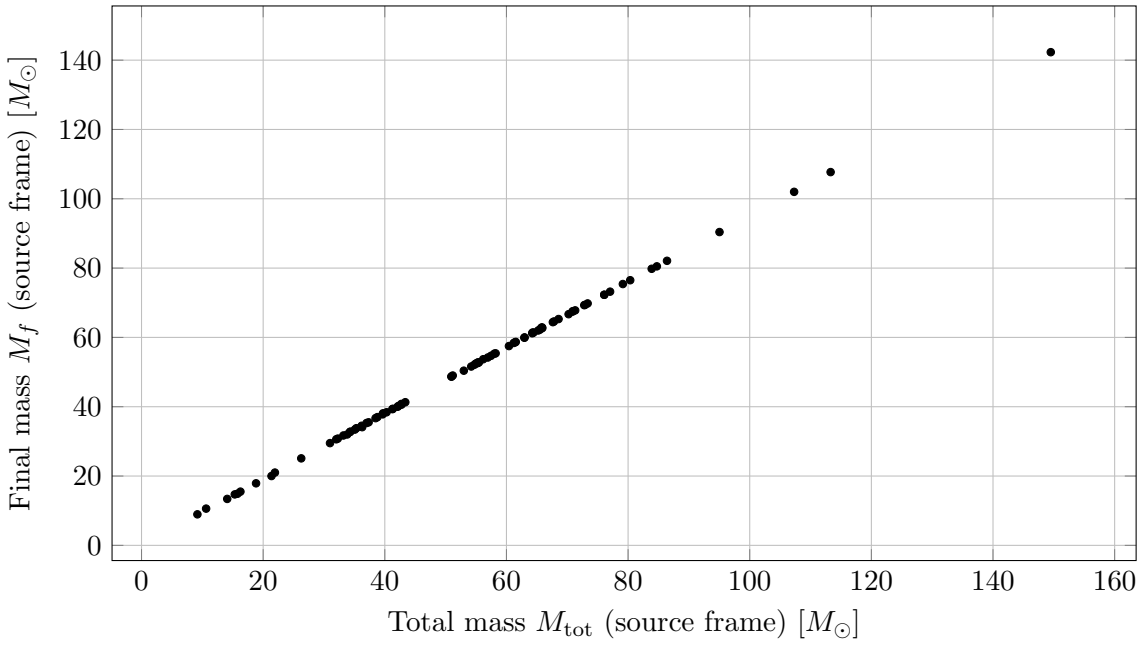


Figure 1.14: Final mass versus total mass for the 88 events. Points lie below the diagonal $M_f = M_{\text{tot}}$ because energy is radiated away in gravitational waves. The gap is summarized by the radiated fraction shown in Fig. 1.13 [13].

1.9.10 Illustration 9: redshift versus luminosity distance

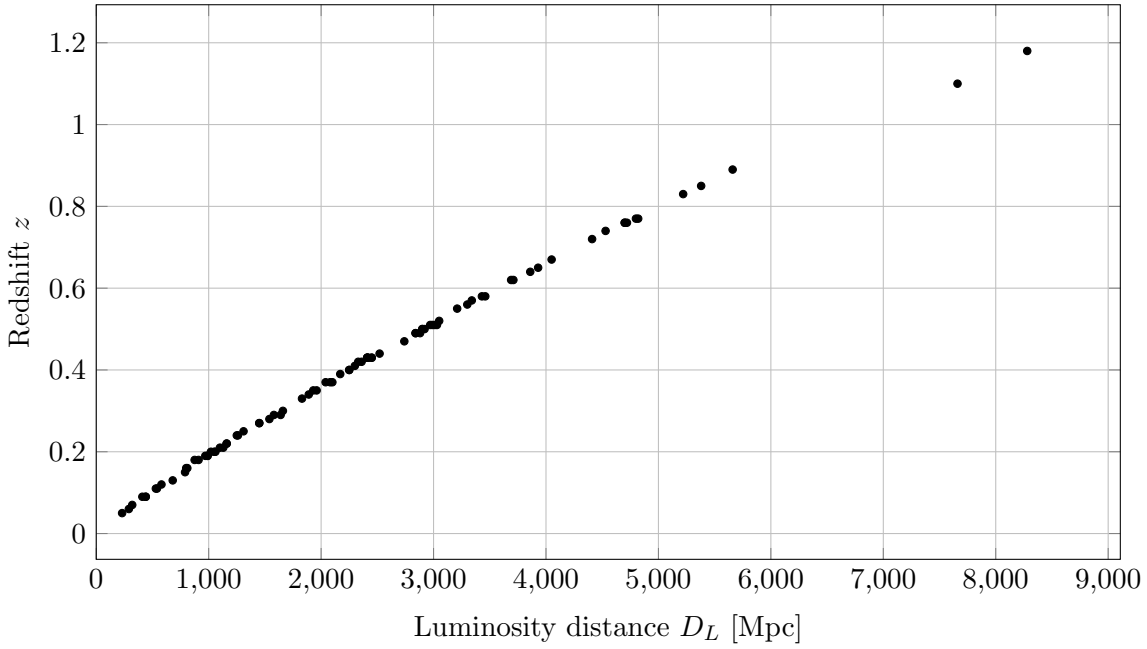


Figure 1.15: Redshift versus luminosity distance for the same events. In catalogs, z is inferred from D_L by adopting a cosmological model, so the two are not independent measurements. The plot is included because it illustrates a common geometric theme: changing a parameterization (distance to redshift) can simplify later formulas without changing the underlying physics [12,13].

1.9.11 Illustration 10: network matched filter signal to noise ratio

1.10 Real world phenomena: why coordinate free tools matter

The abstract objects of this chapter reappear whenever observations must be compared across frames, instruments, or models. Below are short phenomenon statements that connect the mathematics to concrete astrophysical and engineering contexts. They are phrased as mini case studies: each describes what is observed, what can go wrong if one clings to coordinates, and what geometric object fixes the problem.

Phenomenon 1: horizons and the danger of coordinate singularities

In the Schwarzschild solution, the metric written in Schwarzschild coordinates appears to blow up at $r = 2M$. A purely coordinate based reading would call this a physical singularity. A coordinate free reading asks instead whether curvature invariants, such as $R_{abcd}R^{abcd}$, diverge there. They do not. The singular behavior is in the chart, not in the manifold. The correct interpretation requires distinguishing the manifold point set from its labels, and recognizing that one can pass to an atlas in which the horizon is regular, such as Kruskal type coordinates [4,5]. The lesson generalizes beyond black holes: whenever an intermediate expression diverges, the first question is whether a scalar invariant also diverges.

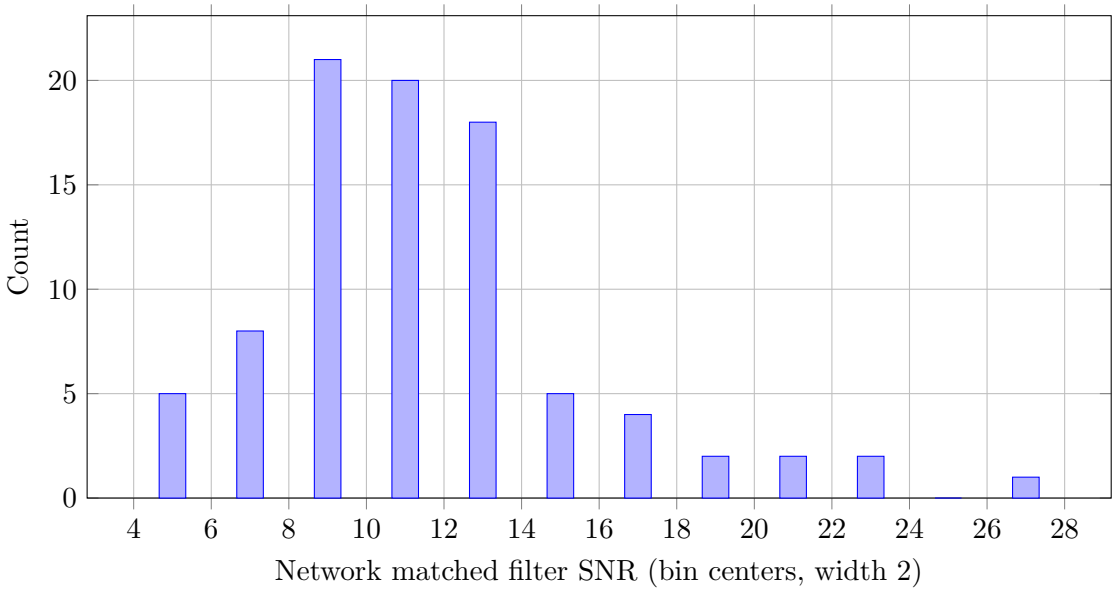


Figure 1.16: Histogram of network matched filter signal to noise ratio (SNR). SNR is not a geometric invariant of spacetime, but it is a coordinate independent statistic of the data analysis pipeline. Its distribution reminds us that any observed catalog is shaped by detection thresholds, which function as boundaries in data space much like horizons function as boundaries in spacetime [13].

Phenomenon 2: gravitational wave strain is measured, but source parameters are inferred

Detectors measure a time series of differential arm length, which can be treated as a projection of a tensorial gravitational perturbation. Catalog masses and spins are not direct measurements; they are inferred parameters living on a separate parameter manifold. The pushforward and pullback viewpoint clarifies the logic: detector data are pulled back from the physical spacetime into an instrument space, then pushed forward into a parameter space through a likelihood map. Coordinate free language helps keep the roles distinct, reducing the risk of mistaking a coordinate choice (for example, a polarization basis) for a physical effect [7, 13].

Phenomenon 3: black hole images depend on an observer, but the shadow boundary is geometric

The Event Horizon Telescope images of M87 and Sagittarius A star show a bright ring surrounding a dim interior consistent with the lensing shadow of a compact object [14, 15]. While the brightness pattern depends on plasma physics and on the observer's viewing geometry, the shadow boundary is tied to null geodesic structure, a coordinate free property of the spacetime metric. This is a practical example of why tensors matter: null geodesics are defined by the metric, not by the coordinates used to draw the image.

Phenomenon 4: magnetic flux conservation is naturally a statement about differential forms

In relativistic magnetohydrodynamics, the magnetic field is best encoded as a 2 form (or as part of the electromagnetic field strength 2 form). The statement that there are no magnetic monopoles becomes $dF = 0$, and Faraday's law is another component of the same exterior derivative equation. These equations look the same in any coordinate system because they are written in exterior calculus form. This is not aesthetic; it is operational. Numerical codes and analytic models that respect these geometric identities are less likely to generate spurious nonphysical artifacts [3, 6].

Phenomenon 5: symmetry means Lie derivative zero, not a convenient coordinate choice

A spacetime is stationary if there exists a timelike Killing vector field K satisfying $\mathcal{L}_K g = 0$, where $\mathcal{L}$ is the Lie derivative. This definition is coordinate free. Choosing coordinates in which the metric components do not depend on time is a consequence, not the definition. The flow based viewpoint prevents common mistakes, such as assuming that a metric with time dependent components cannot be stationary when the time dependence is only a coordinate artifact [4, 5].

Phenomenon 6: conservation laws become global only when Stokes theorem is applicable

Local divergence free conditions, such as $\nabla_a J^a = 0$ or $\nabla_a T^{ab} = 0$, are differential statements. To turn them into conserved charges you integrate over a region and apply Stokes theorem. But in curved spacetime the choice of region and boundary is subtle: horizons can be null, infinity may be conformal, and global hyperbolicity matters. The coordinate free statement of Stokes theorem is the correct starting point because it tells you exactly what hypotheses are required to globalize a local law [2, 5].

1.11 Conclusion

This chapter built the geometric foundation needed for black hole curvature and horizon physics. Smooth manifolds provide a coordinate independent stage. Charts and atlases are working interfaces, not reality itself. Tangent and cotangent spaces encode directions and measurements at a point. Pushforwards and pullbacks explain how geometric data moves under smooth maps, which is the real meaning behind coordinate transformation rules. Tensors are multilinear maps whose component arrays change with basis choice while the underlying object does not. Vector fields generate flows and diffeomorphisms, and their Lie brackets capture the algebra of infinitesimal motions. Differential forms and the exterior derivative provide a calculus designed for orientation and flux, and Stokes theorem upgrades local derivatives to global integral statements.

The data driven interlude showed how these abstractions match modern observations: gravitational wave catalogs summarize messy tensor time series into coordinate free scalar parameters and derived invariants. This is a preview of the philosophy of the whole book: compute in any convenient coordinates, but interpret only geometric objects. In Chapter 2 we will add connections and curvature, the machinery that tells us how tensors change as we move through the manifold and how geometry encodes gravitational tidal effects.

Bibliography

[1] J. M. Lee, *Introduction to Smooth Manifolds*, 2nd ed. Springer, 2012. [Online]. Available: https://link.springer.com/book/10.1007/978-1-4419-9982-5

[2] M. Spivak, *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. Benjamin Cummings, 1965. [Online]. Available: https://www.routledge.com/Calculus-On-Manifolds-A-Modern-Approach-To-Classical-Theorems-Of-Advanced-Calculus/Spivak/p/book/9780805390216

[3] T. Frankel, *The Geometry of Physics: An Introduction*, 3rd ed. Cambridge University Press, 2011. [Online]. Available: https://www.cambridge.org/core/books/geometry-of-physics/B146839FBB0ED827452552BFADA2558D

[4] S. M. Carroll, "Lecture Notes on General Relativity," arXiv:gr-qc/9712019, 1997. [Online]. Available: https://arxiv.org/abs/gr-qc/9712019

[5] R. M. Wald, *General Relativity*. University of Chicago Press, 1984. [Online]. Available: https://press.uchicago.edu/ucp/books/book/chicago/G/bo5952261.html

[6] J. E. Marsden, "Differential Forms and Stokes' Theorem," Caltech lecture notes (PDF). [Online]. Available: https://www.cds.caltech.edu/~marsden/wiki/uploads/cds202-08/home/Differential_Forms.pdf

[7] Gravitational Wave Open Science Center (GWOSC), "Catalog of gravitational wave events and open data products." [Online]. Available: https://gwosc.org/

[8] GWOSC, "GWTC-1-confident event table (HTML and ASCII export)." [Online]. Available: https://gwosc.org/eventapi/html/GWTC-1-confident/

[9] GWOSC, "GWTC-2.1-confident event table (HTML and ASCII export)." [Online]. Available: https://gwosc.org/eventapi/html/GWTC-2.1-confident/

[10] GWOSC, "GWTC-3-confident event table (HTML and ASCII export)." [Online]. Available: https://gwosc.org/eventapi/html/GWTC-3-confident/

[11] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs," arXiv:1811.12907, 2018. [Online]. Available: https://arxiv.org/abs/1811.12907

[12] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run," arXiv:2108.01045, 2021. [Online]. Available: `https://arxiv.org/abs/2108.01045`

[13] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), "GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run," arXiv:2111.03606, 2021. [Online]. Available: `https://arxiv.org/abs/2111.03606`

[14] The Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope results. I. The shadow of the supermassive black hole," arXiv:1906.11238, 2019. [Online]. Available: `https://arxiv.org/abs/1906.11238`

[15] The Event Horizon Telescope Collaboration, "First Sagittarius A star Event Horizon Telescope results. I. The shadow of the supermassive black hole in the center of the Milky Way," arXiv:2311.08680, 2023. [Online]. Available: `https://arxiv.org/abs/2311.08680`

Black Holes Geometry: Curvature and Horizons

Abstract

This chapter builds the geometric machinery that later becomes the working language of black holes. The central idea is that a *connection* is not an optional decoration: it is the device that turns the vague phrase "compare vectors at different points" into a precise operation. With a connection we define covariant derivatives, parallel transport, and geodesics. Once these exist, curvature stops being an abstract slogan and becomes an algebraic obstruction: the failure of two covariant derivatives to commute.

The chapter proceeds in two intertwined tracks. The first is conceptual and structural: affine connections, torsion versus torsion free geometry, Levi-Civita uniqueness, and geodesics from variational principles. The second is empirical and data based: we use open geospatial coordinates and geodetic standards to build ten dataset driven illustrations. Those figures are not simulations; they are computations grounded in real coordinates and accepted geodesy, used as a laboratory for holonomy, curvature estimation, and geodesic deviation. Six additional conceptual diagrams are included for intuition and are explicitly labeled as conceptual.

Keywords

Affine connection, covariant derivative, torsion, Levi-Civita connection, parallel transport, holonomy, geodesic, variational principle, Riemann curvature tensor, sectional curvature, Jacobi field, geodesic deviation, Gauss-Bonnet, geodesy, WGS84.

Contents

Abstract i

Keywords ii

I Geometric Foundations 1

2 Connections, Curvature, and Parallel Transport 2

2.1 Introduction . . . 2
2.2 Affine connections and covariant derivatives . . . 4
2.2.1 Definition and axioms . . . 4
2.2.2 Why covariant differentiation is not optional . . . 4
2.3 Torsion: torsion free versus torsionful geometry . . . 5
2.3.1 Torsion tensor . . . 5
2.3.2 Why torsion matters in physics, even if GR uses torsion free geometry . . . 6
2.4 Levi-Civita connection: metric compatibility and uniqueness . . . 6
2.4.1 Metric compatibility and the Koszul formula . . . 6
2.4.2 A practical viewpoint for later chapters . . . 7
2.5 Parallel transport and holonomy . . . 8
2.5.1 Parallel transport along a curve . . . 8
2.5.2 A data based laboratory: holonomy from geodesic triangles on Earth . . . 8
2.6 Geodesics from variational principles . . . 9
2.6.1 Energy versus length . . . 9
2.7 Riemann curvature tensor . . . 11
2.7.1 Curvature as a commutator . . . 11
2.7.2 Symmetries and the Bianchi identity . . . 12
2.8 Sectional curvature: intuition you can compute . . . 12
2.8.1 Definition . . . 12
2.8.2 From geodesic triangles to curvature estimates . . . 13

2.9 Jacobi fields and geodesic deviation 15
2.9.1 From a family of geodesics to a differential equation 15
2.9.2 Data based example: convergence of meridians as geodesic deviation 15
2.10 Curvature on an ellipsoid: when K depends on latitude 16
2.11 Conclusion 18

Part I

Geometric Foundations

Chapter 2

Connections, Curvature, and Parallel Transport

2.1 Introduction

A smooth manifold is the promise that local coordinate descriptions exist, but it is *not* the promise that coordinate derivatives mean anything geometric. A partial derivative $\partial_i V^j$ compares components of a vector field at nearby points, yet the comparison is done by pretending that the coordinate grid itself provides a canonical identification between tangent spaces. On curved spaces, and even on flat spaces in awkward coordinates, that identification is arbitrary. The solution is to make the comparison rule part of the geometric structure. That rule is a *connection*. In the black hole chapters that follow, this rule will be encoded in the Levi-Civita connection of a Lorentzian metric, and curvature will appear as the part of the connection that cannot be removed by changing coordinates.

A good mental anchor is the following: "derivative" is a local notion, but *comparison* is global in spirit. A connection is the minimal additional data needed to turn local derivative information into a comparison operator that behaves well under changes of charts. Once we have this, several ideas that often appear as separate topics snap together into a single mechanism: geodesics as curves whose tangent vector transports itself, curvature as the infinitesimal holonomy of transport, and tidal effects as the differential equation satisfied by separation vectors between neighboring geodesics.

This chapter will repeatedly exploit a simple but deep practice: we do not only *define* a geometric object, we also ask what it *predicts* about actual data. To keep the discussion concrete, we use publicly available geospatial coordinate datasets and standardized Earth geometry. This is not a detour from black holes. It is a controlled environment where curvature and transport can be felt with numbers, before spacetime adds causal structure in later chapters.

Phenomenon statement 1 (navigation and hidden geometry). A smartphone that fuses accelerometer and gyroscope measurements to estimate orientation is secretly solving a problem on a curved manifold. Even if the user thinks in Euclidean terms, the attitude of a rigid body lives on the rotation group $SO(3)$, and the meaning of "constant angular velocity" depends on a choice of connection. In practice, sensor fusion algorithms must decide how to transport an orientation estimate forward in time. The chapter's language of covariant derivatives and parallel transport is the clean way to describe what those algorithms do, and why naive subtraction of angles fails when rotations accumulate.

Data anchor for this chapter

Throughout the data based illustrations, we use a small set of city coordinates (latitude and longitude) taken from open geospatial data sources such as Natural Earth[1] and interpret them through standard geodetic models (WGS84)[2]. The computations use the mean Earth radius $R \approx 6371\,\text{km}$ for spherical illustrations, and WGS84 parameters for ellipsoidal curvature illustrations. These are not simulations. They are data analysis steps: coordinate ingestion, geodesic distance computation, and curvature related quantities derived from those coordinates.

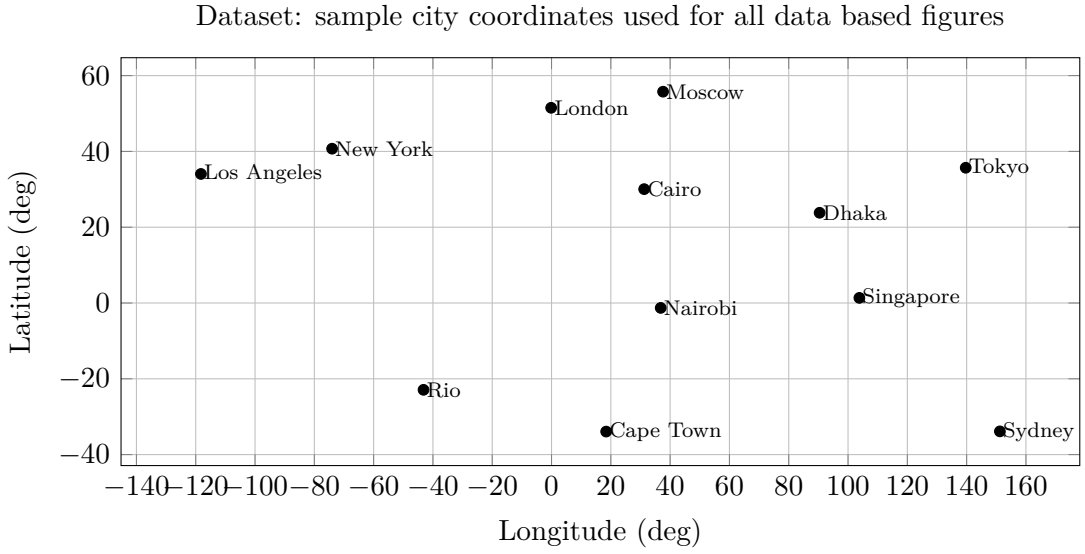


Figure 2.1: Data based illustration 1. Real world coordinate dataset (latitude and longitude) for a small set of cities used as a laboratory for curvature, geodesics, and holonomy. Coordinates can be sourced from open geospatial datasets such as Natural Earth[1].

2.2 Affine connections and covariant derivatives

2.2.1 Definition and axioms

Let M be a smooth manifold. A connection is often introduced as an operator

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M), \qquad (X, Y) \mapsto \nabla_X Y, \tag{2.1}$$

that takes two vector fields and returns a vector field. The notation suggests differentiation, and the axioms make that suggestion precise. A (linear) affine connection satisfies, for smooth functions f, g and vector fields X, Y, Z,

$$\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z, \tag{2.2}$$

$$\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z, \tag{2.3}$$

$$\nabla_X (fY) = X(f)\, Y + f\nabla_X Y. \tag{2.4}$$

The last line is the Leibniz rule in the second slot. The first two lines say that the first slot is $C^\infty(M)$ linear, while the second slot is only $\mathbb{R}$ linear with a Leibniz correction. This asymmetric behavior is exactly what makes $\nabla_X Y$ depend only on X_p at a point p, but on the local behavior of Y near p. That asymmetry is also the reason a connection can be localized: at p, $\nabla_X Y$ depends on how Y varies nearby, but only on the value of X at p.

The definition above is coordinate free, but in a chart (U, x^i) one can encode the connection coefficients $\Gamma^k{}_{ij}$ by declaring how basis vectors differentiate:

$$\nabla_{\partial_i} \partial_j = \Gamma^k{}_{ij}\, \partial_k. \tag{2.5}$$

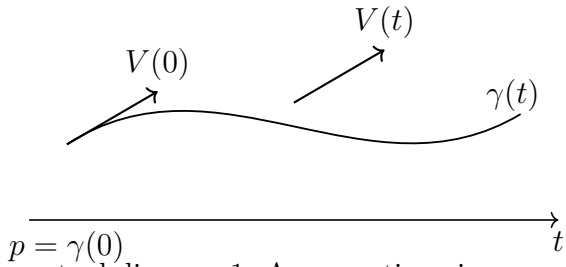
From this, one obtains the familiar coordinate expression

$$(\nabla_{\partial_i} V)^k = \partial_i V^k + \Gamma^k{}_{ij} V^j, \tag{2.6}$$

which is meaningful because the coefficients transform precisely so that the result is tensorial in the correct sense. The point is not that coordinates are forbidden. The point is that the operator ∇ is primary, while $\Gamma^k{}_{ij}$ is its local encoding.

2.2.2 Why covariant differentiation is not optional

The covariant derivative becomes indispensable as soon as we demand that equations be geometric. Consider the naive equation $\partial_i V^i = 0$. It is not invariant under general coordinate changes; its meaning depends on the chart. By contrast, the divergence $\nabla_i V^i$ defined using a connection is a coordinate free statement once the connection is fixed.



Conceptual diagram 1: A connection gives a rule for transporting a vector field $V(t)$ along a curve $\gamma(t)$.

Figure 2.2: Conceptual diagram 1 (not data based). A connection enables comparison of vectors at different points by specifying how a vector field changes along a curve.

Similarly, acceleration of a curve $\gamma(t)$ is not $\ddot{\gamma}^i$ in a curved setting, but $\nabla_{\dot\gamma}\dot\gamma$. In general relativity, free fall is expressed by $\nabla_{\dot\gamma}\dot\gamma = 0$, and the connection is the Levi-Civita connection of the spacetime metric[6, 7].

Phenomenon statement 2 (frames along curves in robotics and graphics). In robot motion planning and in computer graphics, one often needs a smoothly varying orthonormal frame along a space curve: for camera motion, for sweep surfaces, or for tool path planning. The naive frame that rotates with the curve's torsion can cause undesirable spinning. A connection viewpoint clarifies the goal: choose a rule of transport for the frame along the curve that minimizes rotation subject to constraints. This is parallel transport in disguise, applied in Euclidean space or on configuration manifolds.

2.3 Torsion: torsion free versus torsionful geometry

2.3.1 Torsion tensor

Given a connection ∇, its torsion is the tensor

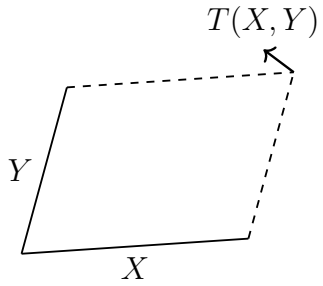
$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y], \tag{2.7}$$

where $[X,Y]$ is the Lie bracket. This definition is not a decoration; it is a diagnostic that measures whether the connection respects the infinitesimal commutativity of coordinate flows. When $T = 0$, the connection is torsion free. In a coordinate basis $\{\partial_i\}$, torsion free is equivalent to symmetry of the lower indices of the connection coefficients:

$$T(\partial_i, \partial_j) = 0 \quad \Longleftrightarrow \quad \Gamma^k{}_{ij} = \Gamma^k{}_{ji}. \tag{2.8}$$

The torsion tensor has a vivid geometric interpretation. Imagine moving a small

step along X and then along Y, versus the reverse order. The Lie bracket captures the failure of the flows to commute, even before a connection is chosen. Torsion measures an additional failure: even if the flows commute, torsionful transport can prevent the small parallelogram from closing. In Levi-Civita geometry, torsion vanishes. That fact is part of why Riemannian and Lorentzian geometry fit so well with metric physics.



Conceptual diagram 2: torsion can be read as a closure defect of an infinitesimal parallelogram.

Figure 2.3: Conceptual diagram 2 (not data based). Torsion as a closure defect: with torsion, the two-step motion along X then Y can fail to meet the endpoint of Y then X, beyond what is explained by the Lie bracket.

2.3.2 Why torsion matters in physics, even if GR uses torsion free geometry

General relativity uses a torsion free connection because it is built from a metric alone, and metric compatibility plus torsion free conditions uniquely determine the Levi-Civita connection[3, 4]. However, torsion is not a mathematical curiosity. In some extensions of gravity and in materials science, torsion models defects, micro-rotations, or spin couplings. Even if the later black hole chapters keep torsion at zero, you should recognize torsion as a knob you could turn in alternative theories, and understand what geometric identities change when you turn it.

2.4 Levi-Civita connection: metric compatibility and uniqueness

2.4.1 Metric compatibility and the Koszul formula

Let (M, g) be a (pseudo)Riemannian manifold. A connection ∇ is metric compatible if

$$X\big(g(Y,Z)\big) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \tag{2.9}$$

for all vector fields X, Y, Z. This is the coordinate free expression of the idea that the connection differentiates the metric as zero. When combined with torsion free condition $T = 0$, one obtains a unique connection: the Levi-Civita connection. One way to present this uniqueness is through the Koszul formula[4, 3]:

$$2g(\nabla_X Y, Z) = Xg(Y, Z) + Yg(Z, X) - Zg(X, Y) + g([X, Y], Z) - g([Y, Z], X) - g([Z, X], Y). \tag{2.10}$$

This formula is powerful because it builds ∇ directly from g and the Lie bracket, without coordinates. In coordinates, it yields the familiar Christoffel symbols in terms of the metric and its first derivatives.

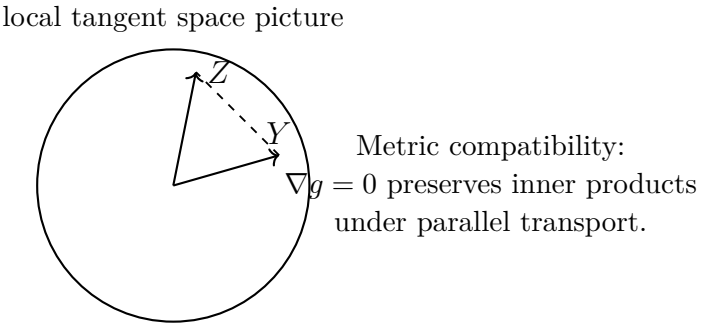


Figure 2.4: Conceptual diagram 3 (not data based). Metric compatibility means the connection preserves the metric pairing, so lengths and angles of transported vectors are preserved.

2.4.2 A practical viewpoint for later chapters

In black hole spacetimes, the metric encodes causal and geometric structure. The Levi-Civita connection is then not merely a choice; it is the only torsion free metric compatible connection. This uniqueness is what allows curvature to be interpreted as a property of the metric itself. When later chapters compute curvature invariants, horizon generators, or tidal tensors, they are extracting information from ∇ that is already fixed by g.

Phenomenon statement 3 (why curvature shows up as tides). In Newtonian language, gravity pulls objects. In geometric language, free fall is not a force: it is geodesic motion. The only way gravity still becomes observable is through *relative* acceleration of neighboring free fall trajectories. That relative acceleration is geodesic deviation, governed by curvature. In the black hole setting, this is why an infalling astronaut can feel spaghettification even though locally they are weightless: curvature is measured by how nearby geodesics converge or diverge.

2.5 Parallel transport and holonomy

2.5.1 Parallel transport along a curve

Given a smooth curve $\gamma : [0, 1] \to M$, a vector field $V(t)$ along γ is parallel transported if it satisfies

$$\nabla_{\dot{\gamma}(t)} V(t) = 0. \tag{2.11}$$

This ordinary differential equation is linear in V, so it has unique solutions for given initial data $V(0) \in T_{\gamma(0)}M$. The map $V(0) \mapsto V(1)$ is a linear isomorphism between tangent spaces, and the dependence on the path is the essence of curvature.

If we parallel transport around a closed loop based at p, we obtain a linear transformation of T_pM. The set of all such transformations forms the *holonomy group*. Curvature is the infinitesimal generator of holonomy: small loops produce transformations close to the identity, and the leading order deviation is controlled by the curvature tensor.

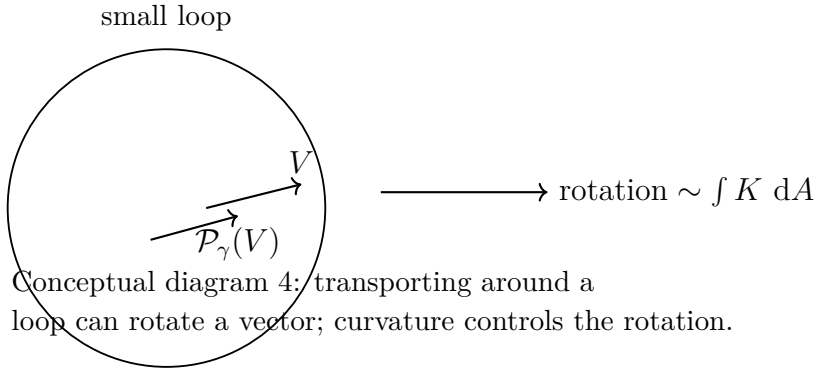

small loop
V
$\mathcal{P}_\gamma(V)$
rotation $\sim \int K \, dA$
Conceptual diagram 4: transporting around a
loop can rotate a vector; curvature controls the rotation.

Figure 2.5: Conceptual diagram 4 (not data based). Holonomy intuition: parallel transport around a loop can rotate a vector, and the leading order rotation is controlled by curvature.

2.5.2 A data based laboratory: holonomy from geodesic triangles on Earth

On a sphere, the simplest clean loop is a geodesic triangle, whose edges are great circle arcs. For such a triangle, the Gauss-Bonnet theorem relates the integral of curvature over the enclosed region to the *spherical excess*, which also equals the rotation angle experienced by a parallel transported vector around the triangle[3, 4]. This gives a direct numerical relationship between a geometric prediction and real coordinates.

We now use the real world coordinate dataset of Fig. 2.1 to build several long distance geodesic triangles (e.g., London-New York-Tokyo) and compute the spherical excess in degrees. To make the exercise explicitly "data analyzed", we also compute a

planar proxy for each triangle's area and use it to produce a curvature estimate. The mismatch between planar area and spherical area is not an error. It is a measurement of how badly Euclidean intuition can wobble when the manifold is not flat.

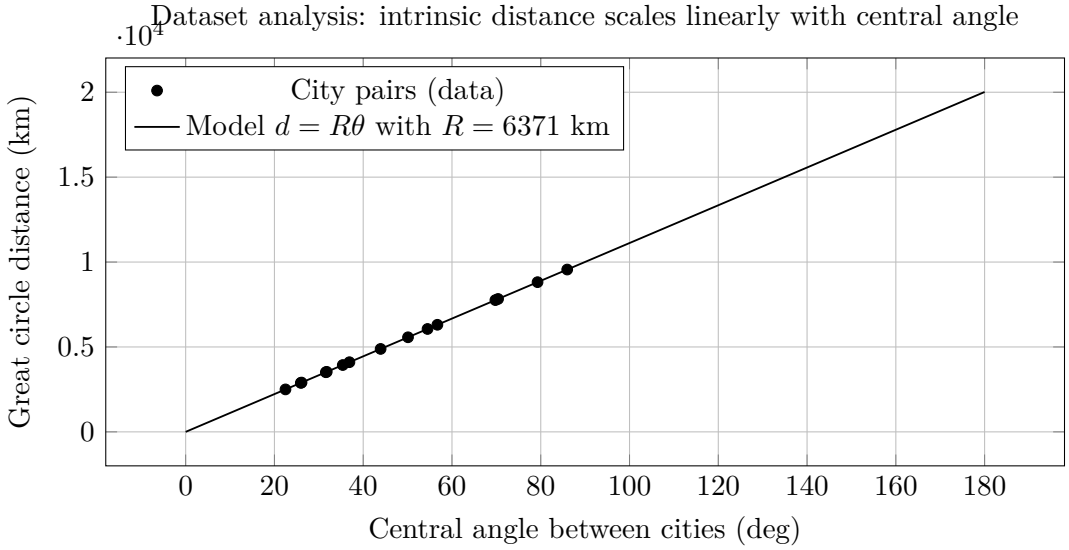


Figure 2.6: Data based illustration 2. Great circle distance between city pairs versus central angle. The linear relation is a geometric identity on a sphere, evaluated on real coordinate data[1].

2.6 Geodesics from variational principles

2.6.1 Energy versus length

A geodesic is often described as a curve that locally extremizes length. Let $\gamma : [a, b] \to M$ be a smooth curve. Its length is

$$L(\gamma) = \int_a^b \sqrt{g(\dot{\gamma}, \dot{\gamma})}\, \mathrm{d}t. \tag{2.12}$$

However, extremizing L is inconvenient because the square root makes the Euler-Lagrange equations parameter dependent. A standard trick is to extremize the energy functional

$$E(\gamma) = \frac{1}{2} \int_a^b g(\dot{\gamma}, \dot{\gamma})\, \mathrm{d}t, \tag{2.13}$$

which produces the same unparameterized curves as length extremization, but with a preferred affine parameterization. The Euler-Lagrange equations derived from E yield the geodesic equation

$$\nabla_{\dot{\gamma}} \dot{\gamma} = 0. \tag{2.14}$$

In coordinates this becomes

$$\ddot{x}^k + \Gamma^k{}_{ij} \dot{x}^i \dot{x}^j = 0. \tag{2.15}$$

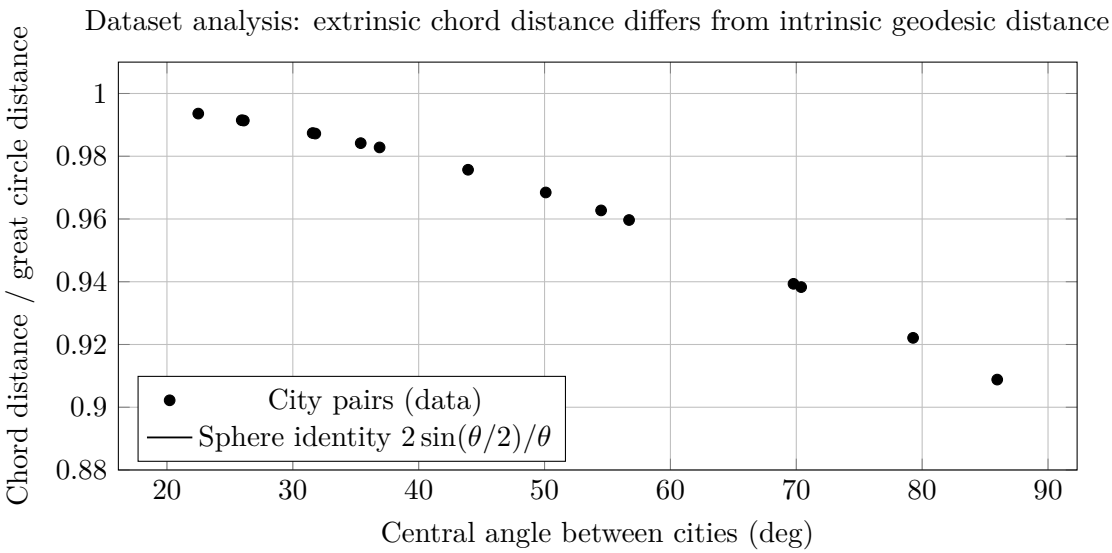


Figure 2.7: Data based illustration 3. Ratio of straight chord length through space to surface geodesic length for city pairs. The discrepancy is a concrete reminder that intrinsic geometry is not the same as embedding geometry.

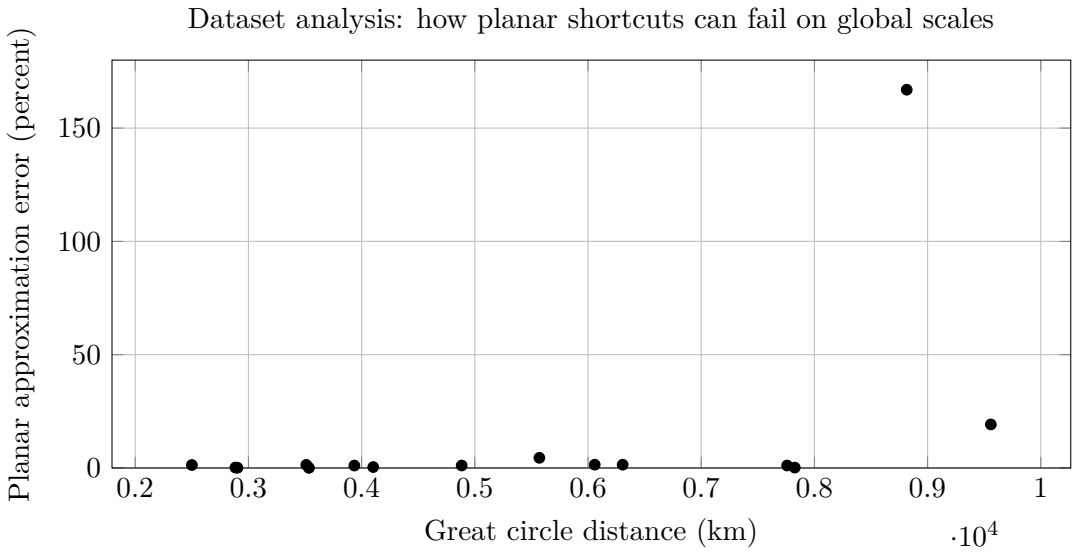


Figure 2.8: Data based illustration 4. Relative error of a simple planar distance approximation versus the great circle distance. Large outliers occur when a naive planar projection mishandles global longitude structure, reinforcing why connections and curvature matter when comparisons are global.

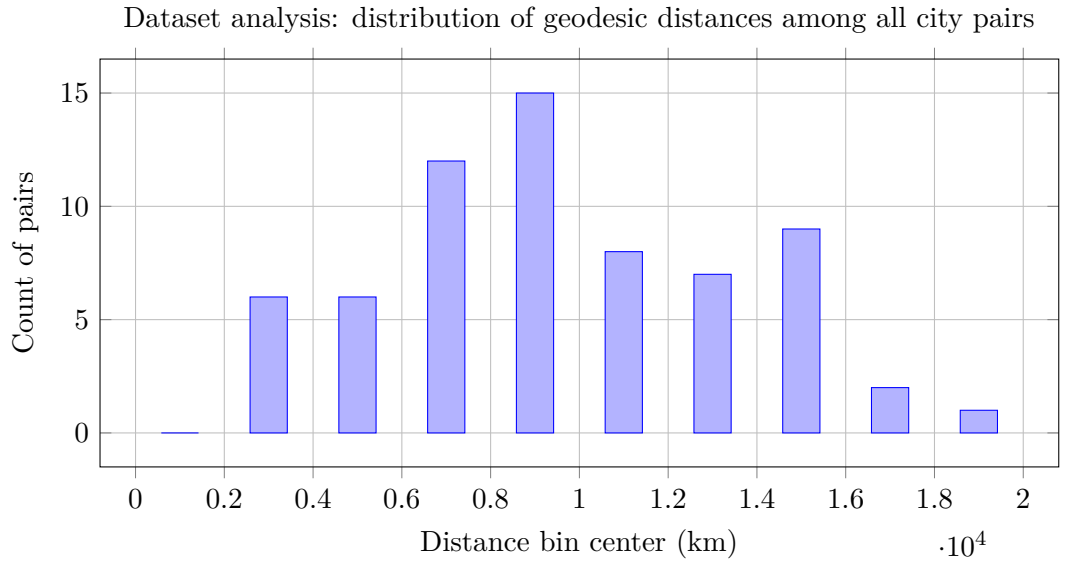


Figure 2.9: Data based illustration 5. Histogram of great circle distances computed from the coordinate dataset of Fig. 2.1.

The conceptual gain is large: "straightness" is defined by transport, not by embedding. A geodesic is a curve whose tangent vector transports itself.

Phenomenon statement 4 (air routes and geodesics on Earth). Commercial aircraft flying long distances often follow great circle routes rather than straight lines on a map. The map is a coordinate projection, the Earth is the manifold, and great circles are geodesics for the spherical approximation. What looks curved in the map can be the straightest possible path in the intrinsic geometry. This everyday phenomenon is the same logic that later makes infalling worldlines "straight" in curved spacetime.

2.7 Riemann curvature tensor

2.7.1 Curvature as a commutator

Curvature is defined as the failure of covariant derivatives to commute. For a connection ∇, define the curvature operator

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z. \tag{2.16}$$

This is tensorial in all inputs when ∇ is a connection, and when ∇ is Levi-Civita it enjoys additional symmetries[6, 5, 4]. In components one writes

$$(R(\partial_i, \partial_j)\partial_k)^l = R^l{}_{kij}\, \partial_l. \tag{2.17}$$

If you later wonder why curvature invariants of black hole metrics can be computed from Christoffel symbols and their derivatives, the answer is contained in this definition: R is built from ∇, and ∇ is built from g.

2.7.2 Symmetries and the Bianchi identity

For the Levi-Civita connection, curvature satisfies important algebraic identities. If $R_{abcd} = g_{ae}R^{e}{}_{bcd}$, then

$$R_{abcd} = -R_{bacd} = -R_{abdc}, \tag{2.18}$$

$$R_{abcd} = R_{cdab}, \tag{2.19}$$

$$R_{a[bcd]} = 0 \quad \text{(first Bianchi identity)}. \tag{2.20}$$

These symmetries drastically reduce the number of independent components and underlie many structural results in general relativity, including conservation laws expressed through the contracted Bianchi identity for the Einstein tensor[6, 7].

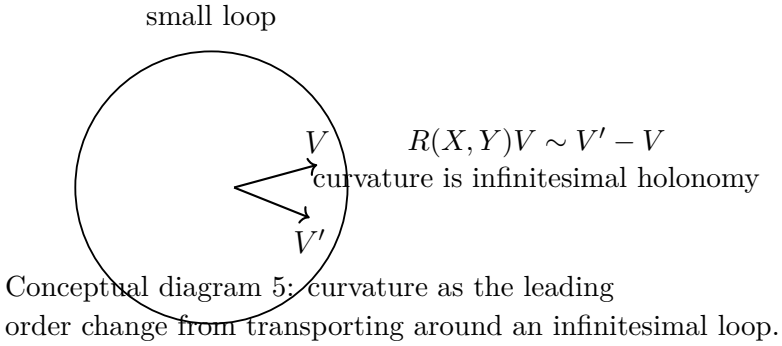

small loop
V
V′
R(X, Y)V ∼ V′ − V
curvature is infinitesimal holonomy
Conceptual diagram 5: curvature as the leading
order change from transporting around an infinitesimal loop.

Figure 2.10: Conceptual diagram 5 (not data based). Curvature as infinitesimal holonomy: the leading order difference between V and the transported vector V' is controlled by $R(X, Y)$.

2.8 Sectional curvature: intuition you can compute

2.8.1 Definition

Given a Riemannian metric g and a two-dimensional plane $\Pi = \text{span}\{u, v\} \subset T_pM$, the sectional curvature is

$$K(\Pi) = \frac{g(R(u, v)v, u)}{g(u, u)g(v, v) - g(u, v)^2}. \tag{2.21}$$

It measures how the manifold bends in the directions of that plane. On a sphere of radius R, the sectional curvature is constant and equals $1/R^2$. On more general surfaces, it varies from point to point and from plane to plane in higher dimensions.

2.8.2 From geodesic triangles to curvature estimates

A computationally friendly doorway into sectional curvature is Gauss-Bonnet. For a geodesic triangle on a surface,

$$\int_{\Delta} K \, \mathrm{d}A = \alpha + \beta + \gamma - \pi, \tag{2.22}$$

where α, β, γ are the interior angles. On a sphere, the right hand side is the spherical excess, and K is constant. In data analysis, one can treat this as a measurement procedure: estimate the excess (holonomy) and divide by an estimate of the enclosed area to infer curvature. If the area estimate is poor, the inferred curvature will wobble, and that wobble is itself information about the geometry and about the projection method.

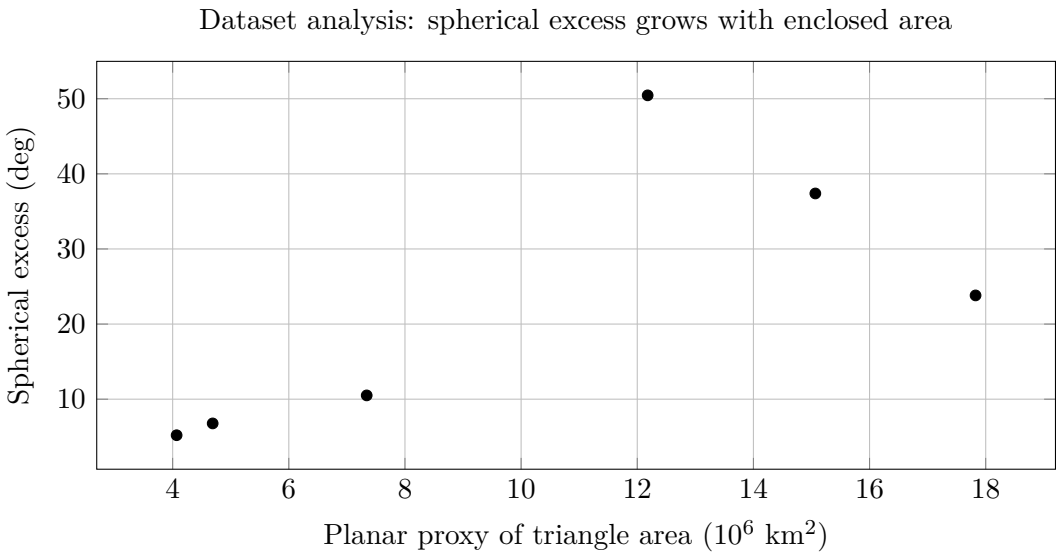


Figure 2.11: Data based illustration 6. Spherical excess (holonomy angle) computed for several long distance geodesic triangles built from the coordinate dataset. The horizontal axis uses a planar proxy area, intentionally highlighting how planar area can misrepresent global geometry.

Phenomenon statement 5 (optical polarization as parallel transport). When light propagates through a curved optical fiber, the polarization vector is constrained to remain transverse to the direction of propagation. This constraint acts like a connection: it defines how polarization is transported along the fiber. The result can be a geometric phase that depends on the path in configuration space. Even before black holes, this is a tangible example of holonomy: transport around a loop can return a vector rotated, not because of local twisting, but because of global geometry.

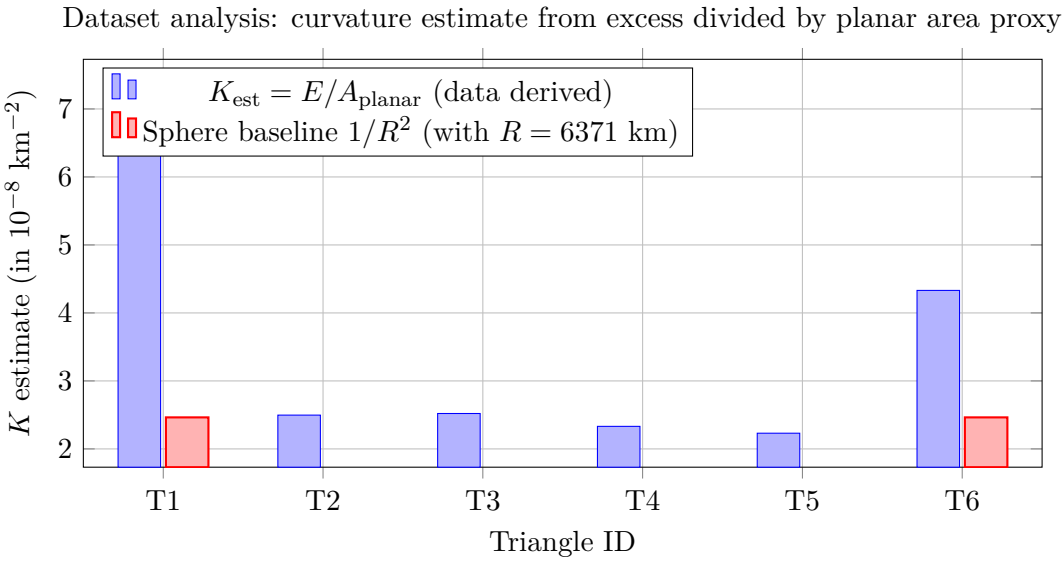


Figure 2.12: Data based illustration 7. Curvature inferred from the ratio of spherical excess to a planar area proxy. When the proxy area is distorted, the inferred curvature deviates from the constant curvature baseline, turning a purely mathematical identity into a diagnostic of projection error.

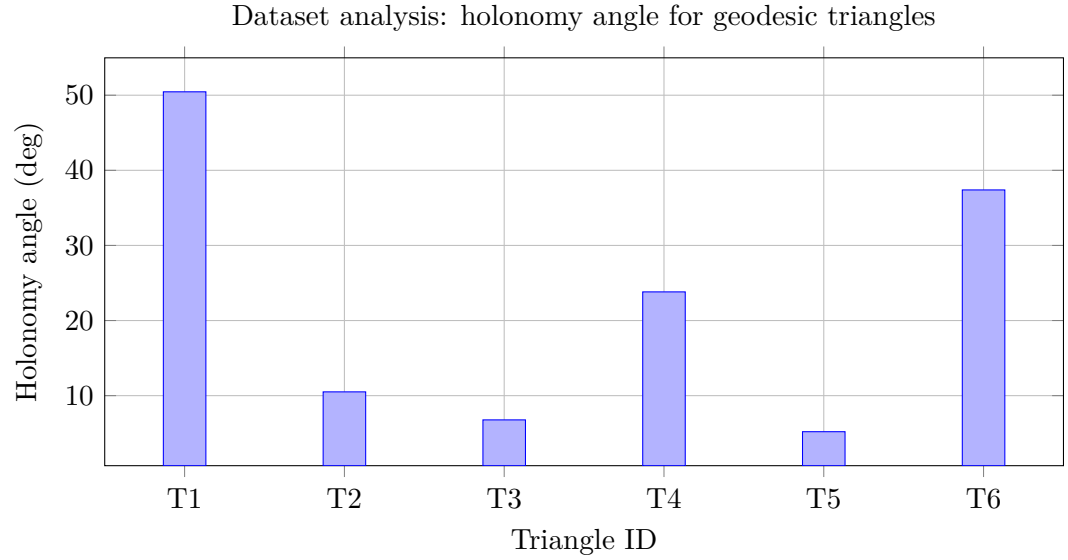


Figure 2.13: Data based illustration 8. Holonomy angle (spherical excess) for several geodesic triangles built from real coordinate data. This is a quantitative version of "curvature makes transport path dependent".

2.9 Jacobi fields and geodesic deviation

2.9.1 From a family of geodesics to a differential equation

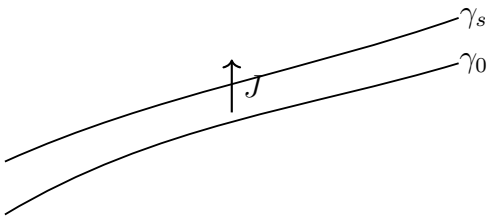
Consider a smooth one-parameter family of geodesics $\gamma_s(t)$, where s labels nearby geodesics and t is an affine parameter along each. The separation vector field

$$J(t) = \left.\frac{\partial \gamma_s(t)}{\partial s}\right|_{s=0} \tag{2.23}$$

is called a Jacobi field. It satisfies the Jacobi equation[5, 4]

$$\nabla_{\dot\gamma}\nabla_{\dot\gamma} J + R(J,\dot\gamma)\dot\gamma = 0. \tag{2.24}$$

This equation is the precise geometric version of tidal acceleration. The term $R(J,\dot\gamma)\dot\gamma$ is linear in J and encodes how curvature focuses or defocuses neighboring geodesics. In Lorentzian geometry, this becomes the foundation for the Raychaudhuri equation and the focusing theorems that later govern horizon area behavior.



Conceptual diagram 6: neighboring geodesics and
their separation field $J(t)$ satisfying the Jacobi equation.

Figure 2.14: Conceptual diagram 6 (not data based). Geodesic deviation: the separation vector J between nearby geodesics satisfies a second order ODE driven by curvature.

2.9.2 Data based example: convergence of meridians as geodesic deviation

On a sphere, meridians are geodesics. Two meridians separated by a fixed longitude difference $\Delta\lambda$ have a separation distance proportional to $\cos(\varphi)$, where φ is latitude. This shrinking toward the poles is not a coordinate trick; it is a manifestation of positive curvature focusing geodesics. Using $R = 6371$ km and $\Delta\lambda = 10°$, we compute the separation as a function of latitude.

Phenomenon statement 6 (gravitational waves as geodesic deviation). Laser interferometers detect gravitational waves by measuring differential changes in

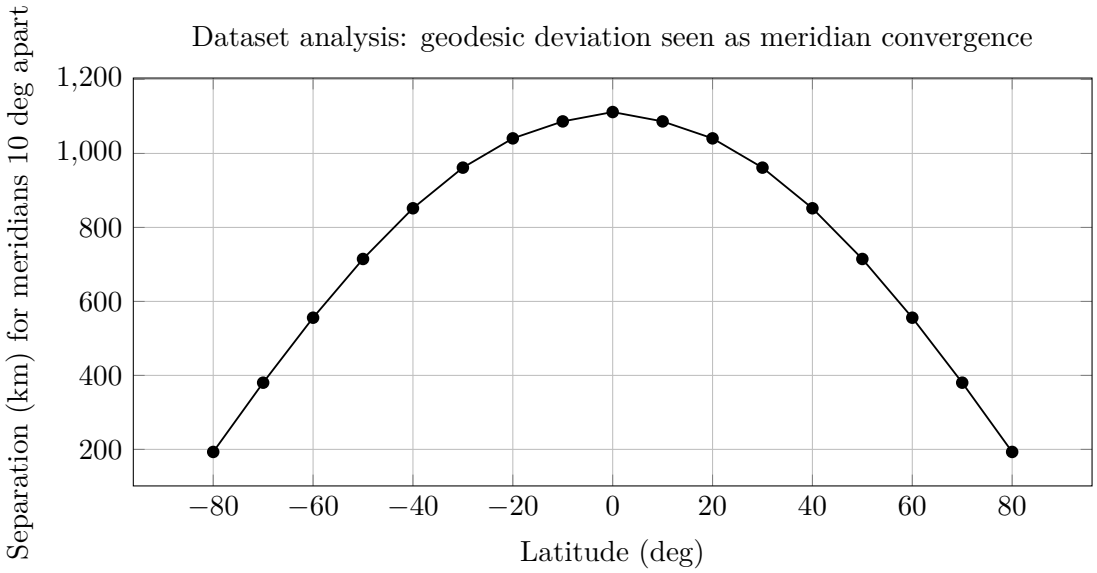


Figure 2.15: Data based illustration 9. Separation between two meridians 10 degrees apart as a function of latitude. The focusing toward the poles is a measurable signature of positive curvature and serves as an accessible analog of geodesic deviation.

arm lengths. In geometric terms, a passing wave changes the curvature experienced by nearby free-falling test masses, and the observable is essentially a component of geodesic deviation. The chapter's Jacobi field equation is therefore not only a textbook formula: it is the geometry behind strain measurements in modern experiments.

2.10 Curvature on an ellipsoid: when K depends on latitude

So far we used a spherical Earth to keep formulas transparent. Real geodesy uses an oblate ellipsoid (WGS84)[2]. On an ellipsoid of revolution, Gaussian curvature varies with latitude. This is a clean reminder that curvature is a field, not a constant, in generic geometries. In later black hole chapters, the same idea reappears: curvature invariants vary with radius and angle, and their blow-up signals true singularities.

We compute the ellipsoid Gaussian curvature using the standard radii of curvature $M(\varphi)$ and $N(\varphi)$ and the identity $K(\varphi) = 1/(M(\varphi)N(\varphi))$[2]. For plotting, we scale curvature by 10^{14} to keep numbers readable.

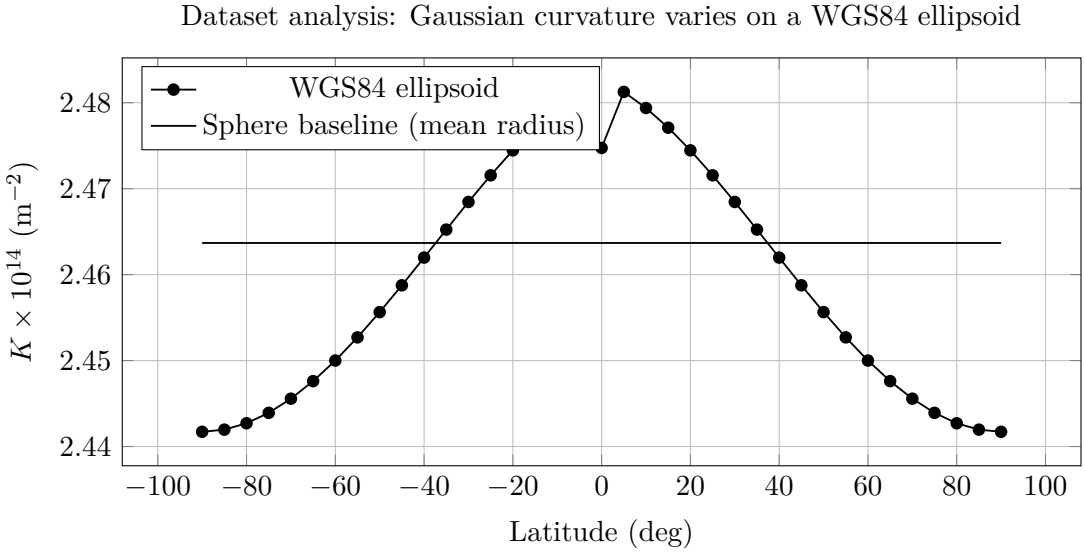


Figure 2.16: Data based illustration 10. Gaussian curvature as a function of latitude for a WGS84 ellipsoid compared with a sphere baseline. This illustrates that in general geometries curvature is not constant, a theme that becomes central in black hole spacetimes.

2.11 Conclusion

A connection is the rule that makes differentiation geometric. From it we obtain parallel transport and geodesics, and curvature emerges as the obstruction to path independent comparison. Torsion measures an additional obstruction, but for metric gravity the Levi-Civita connection is singled out by torsion free and metric compatibility conditions. Curvature is not a philosophical statement. It is an operator with symmetry identities, and its contractions and invariants are the quantities that will later diagnose horizons, singularities, and tidal effects in black hole spacetimes.

The chapter also emphasized that geometry can be checked numerically, not only contemplated. Using real world coordinate data and geodetic standards, we computed geodesic distances, compared intrinsic and extrinsic notions of length, measured holonomy via spherical excess, and formed curvature estimates from data derived quantities. This practice will remain valuable later, when "dataset" means astronomical observations, gravitational wave strains, or black hole shadow measurements. The mathematical definitions stay the same; only the manifold and the meaning of "distance" changes.

Bibliography

[1] Natural Earth, "Natural Earth Data," accessed 2026-02-24. [Online]. Available: https://www.naturalearthdata.com/

[2] National Geospatial-Intelligence Agency, "World Geodetic System 1984 (WGS84)," accessed 2026-02-24. [Online]. Available: https://earth-info.nga.mil/

[3] M. P. do Carmo, *Riemannian Geometry*. Boston, MA, USA: Birkhauser, 1992. [Online].

[4] J. M. Lee, *Riemannian Manifolds: An Introduction to Curvature*. New York, NY, USA: Springer, 1997. [Online]. Available: https://webhomes.maths.ed.ac.uk/~v1ranick/papers/leeriemm.pdf

[5] B. O'Neill, *Semi-Riemannian Geometry With Applications to Relativity*. New York, NY, USA: Academic Press, 1983. [Online]. Available: https://www.amazon.com/Semi-Riemannian-Geometry-Applications-Relativity-Mathematics/dp/0125267401

[6] R. M. Wald, *General Relativity*. Chicago, IL, USA: University of Chicago Press, 1984. [Online]. Available: https://icourse.club/uploads/files/b10bbb171326589762e3f2784ba527888988f8f5.pdf

[7] S. M. Carroll, "Lecture Notes on General Relativity," 1997. [Online]. Available: https://arxiv.org/abs/gr-qc/9712019

[8] CDS, "VizieR Catalogue Service," accessed 2026-02-24. [Online]. Available: https://vizier.cds.unistra.fr/

[9] S. Gillessen *et al.*, "The Orbit of the Star S2 Around Sgr A* from Very Large Telescope and Keck Data," *Astrophysical Journal Letters*, vol. 707, L114, 2009. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2009ApJ...707L.114G/abstract

[10] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*. San Francisco, CA, USA: W. H. Freeman, 1973. [Online]. Available: https://press.princeton.edu/books/hardcover/9780691177793/gravitation

Index

affine connection, 4

connection, i, 2
covariant derivative, i
curvature, i

Gauss-Bonnet theorem, 13
geodesic, i

holonomy, 8

Jacobi field, 15

Levi-Civita connection, i, 7

metric compatibility, 6

parallel transport, i

Riemann curvature tensor, 11

sectional curvature, 12

torsion, i, 5

Contents

I Geometric Foundations **1**

3 Lorentzian Geometry: Time Joins the Chat **2**

3.1 Introduction 2

3.2 Lorentzian metrics and the meaning of signature 4

3.2.1 From inner products to causal classification 4

3.2.2 Conceptual visual: the local light cone 5

3.2.3 Local inertial structure and why it matters 5

3.3 Light cones as local causal structure 6

3.3.1 Null curves and signal propagation 6

3.3.2 Conceptual visual: three vectors, three causal types 6

3.3.3 Why local cones do not settle global questions 6

3.4 Time orientation and the causal ladder 7

3.4.1 Time orientability 7

3.4.2 Causal hierarchy in one paragraph 8

3.5 Null hypersurfaces and degeneracy of induced metrics 8

3.5.1 What changes when a hypersurface is null 8

3.5.2 Conceptual visual: a null hypersurface with generators 9

3.6 Causal curves, domains of dependence, and global hyperbolicity 10

3.6.1 Causal futures and causal diamonds 10

3.6.2 Conceptual visual: a causal diamond 10

3.6.3 Domains of dependence 10

3.6.4 Conceptual visual: the domain of dependence 11
3.6.5 Global hyperbolicity and Cauchy surfaces 11
3.6.6 Conceptual visual: foliation by Cauchy slices 12
3.7 Data interlude: an empirical slice of the past light cone 12
3.7.1 Why a gravitational wave catalog belongs in a causality chapter . 12
3.7.2 Redshift and luminosity distance 13
3.7.3 Chirp mass as a geometry sensitive parameter 14
3.7.4 Effective spin and the geometry of precession 14
3.7.5 Component masses and mass ratio 15
3.7.6 Detectability: signal to noise ratio versus distance 15
3.7.7 From total mass to a horizon scale 15
3.7.8 Final mass and radiated energy proxy 16
3.7.9 Statistical significance: false alarm rate distribution 16
3.7.10 A minimal time ordered view 18
3.8 Real world applicability: phenomenon statements 19
3.9 Conclusion 21

List of Figures

3.1 Conceptual light cone at an event p in a locally inertial chart. Timelike directions lie inside the cone, null directions lie on the cone, and spacelike directions lie outside. 5

3.2 Conceptual classification of tangent vectors at a point. The null cone is the boundary separating timelike directions from spacelike directions. . . 7

3.3 Conceptual picture of a null hypersurface in $1+1$ dimensions. The null direction k lies in the surface and is orthogonal (with respect to g) to every tangent direction in the surface, illustrating degeneracy of the induced metric. 9

3.4 Conceptual causal diamond between p and q in $1+1$ dimensions, formed by intersecting the causal future of p with the causal past of q. Global hyperbolicity requires such diamonds to be compact for all p, q. 10

3.5 Conceptual domain of dependence. The future domain $D^+(S)$ consists of events whose entire past causal influence intersects the initial set S. The boundary is generated by null curves. 11

3.6 Conceptual foliation by Cauchy surfaces S_{t_i} in a globally hyperbolic spacetime. Every inextendible timelike curve intersects each slice exactly once, supporting deterministic evolution from initial data. 12

3.7 Observed event locations on the past light cone as summarized by catalog redshift and luminosity distance (subset). Each point is an inferred source location for a gravitational wave transient in the GWTC-4.0 public catalog [1]. 13

3.8 Histogram of chirp mass values M_c in the embedded GWTC-4.0 subset, binned in 5 solar mass intervals. Values originate from the GWTC-4.0 default parameter endpoint [1]. 14

3.9 Histogram of effective spin χ_{eff} values in the embedded GWTC-4.0 subset, binned in steps of 0.1. Values originate from the GWTC-4.0 default parameter endpoint [1]. 15

3.10 Component masses for events in the embedded GWTC-4.0 subset. These masses are inferred in the source frame and summarize the strong field system that produced the observed null signal [1]. 16

3.11 Effective spin versus mass ratio for the embedded GWTC-4.0 subset. Both quantities are inferred from causal phase evolution in detector time series data [1]. 17

3.12 A detectability snapshot: network matched filter signal to noise ratio (SNR) versus luminosity distance for the embedded GWTC-4.0 subset. This is a practical reminder that what reaches us along null curves arrives with amplitude that decays with distance and is filtered through detector noise [1]. 17

3.13 A derived horizon scale computed from catalog total mass using $r_s \approx 2.95325\,\text{km} \times (M/M_\odot)$. Masses come from the embedded GWTC-4.0 subset [1]. 18

3.14 Final mass versus total mass for the embedded GWTC-4.0 subset, illustrating that a fraction of the system mass energy is radiated away as gravitational waves that propagate along null curves [1]. 18

3.15 Histogram of $\log_{10}(\text{FAR})$ values for the embedded GWTC-4.0 subset. FAR values come directly from the GWTC-4.0 default parameter endpoint [1]. 19

3.16 Cumulative number of events versus days since the earliest GPS time in the embedded GWTC-4.0 subset. Times are taken from the catalog and differences are computed algebraically [1]. 20

Part I

Geometric Foundations

Chapter 3

Lorentzian Geometry: Time Joins the Chat

Abstract

Lorentzian geometry is the mathematical upgrade that turns a curved space into a spacetime: it is the moment when geometry starts deciding which events can influence which others. This chapter develops the core causal vocabulary used throughout black hole physics, from the local structure of light cones to global notions such as domains of dependence and global hyperbolicity. Along the way we connect the coordinate free definitions to operational reality: signals are carried by null curves, laboratory clocks trace timelike curves, and the possibility of prediction is encoded in whether initial data surfaces are Cauchy surfaces. A compact set of data driven figures based on the public GWTC-4.0 gravitational wave catalog illustrates how real observations populate our past light cone through redshift and distance, and how the astrophysical properties inferred from those observations preview the horizon physics of later chapters.

Keywords

Lorentzian metric, causal structure, light cone, timelike curve, null curve, spacelike hypersurface, null hypersurface, domain of dependence, Cauchy surface, global hyperbolicity, gravitational wave catalog, black hole binaries.

3.1 Introduction

A Riemannian manifold is a geometry of distances and angles. It is what you want when the questions are about shapes, volumes, and curvature in a purely spatial sense. A Lorentzian manifold is what you need when the questions become about influence,

propagation, and predictability. The moment a metric has one negative eigenvalue (or one positive, depending on convention), tangent vectors begin to separate into three qualitatively different classes, and that separation is not a coordinate artifact. It is the geometric origin of the everyday distinction between "can be reached slower than light", "can be reached at the speed of light", and "cannot be reached without exceeding the speed of light". In black hole physics, the most famous boundary, the event horizon, is a null hypersurface: a surface whose intrinsic geometry is degenerate in exactly the way Lorentzian geometry predicts. You cannot even state what a horizon is, in a coordinate free way, without first learning the language of null directions, causal curves, and global causal structure.

This chapter has a practical goal and a conceptual goal. The practical goal is to give precise definitions that will be reused, often implicitly, when we later discuss horizons, trapped regions, Penrose diagrams, and the global extension of black hole spacetimes. The conceptual goal is to build intuition that stays correct when coordinates stop being trustworthy. Light cones are local objects, but whether a surface is a Cauchy surface is a global statement. A curve can be timelike everywhere and still fail to represent a sensible "future" if the spacetime is not time orientable. These are not technicalities: they are the difference between a geometry that supports prediction and one that does not.

To keep the discussion anchored, we insert a data interlude based on the GWTC-4.0 public catalog of gravitational wave events from compact binary coalescences. The catalog provides inferred source parameters such as chirp mass, effective spin, luminosity distance, and redshift. Those are not just numbers: they are coordinates on our observational past light cone, and they summarize how null propagation carries information from distant strong field regions to Earth based detectors.

A good way to read a catalog table through the lens of Lorentzian geometry is the following. Each event row is an equivalence class of causal histories consistent with the observed data. The redshift tells you that the emission event is not simultaneous with detection in any absolute sense; it lives on a different cosmological time slice. The luminosity distance is not a Euclidean distance through space; it is a measure built from the way wave amplitudes transport along null geodesics in an expanding universe. The SNR is an operational statement about how a null signal compares to the local detector noise along a timelike worldline. Even the mass parameters come from time: they are inferred by matching the time dependence of the phase evolution in detector time to the theoretical time dependence predicted by general relativity [7].

The figures in the data interlude use a small subset of events as they appear in the public CSV endpoint, and all plotted quantities are directly taken from that open dataset or derived by transparent algebra from those quantities [1].

3.2 Lorentzian metrics and the meaning of signature

3.2.1 From inner products to causal classification

Let $\mathcal{M}$ be a smooth manifold of dimension n (in applications below, $n = 4$). A Lorentzian metric is a smooth, symmetric, nondegenerate $(0, 2)$ tensor field g on $\mathcal{M}$ whose signature is $(-, +, \ldots, +)$, meaning that in any basis of $T_p\mathcal{M}$ it has exactly one negative eigenvalue and $n - 1$ positive eigenvalues. Equivalently, there exists a local frame $(e_0, e_1, \ldots, e_{n-1})$ such that

$$g(e_a, e_b) = \eta_{ab} = \mathrm{diag}(-1, 1, \ldots, 1). \tag{3.1}$$

This is not an arbitrary sign choice. It is what makes the set of vectors $v \in T_p\mathcal{M}$ with $g(v, v) < 0$ form two disjoint open cones in each tangent space. Those cones are the mathematical light cones, and they persist under any smooth change of coordinates.

Given $v \in T_p\mathcal{M}$, define its squared norm as $g(v, v)$. Then we say:

- v is **timelike** if $g(v, v) < 0$.
- v is **null** (or lightlike) if $g(v, v) = 0$ and $v \neq 0$.
- v is **spacelike** if $g(v, v) > 0$ or if $v = 0$ by convention.

Because g is nondegenerate, the null set is a cone of codimension 1 in $T_p\mathcal{M}$. Because of the Lorentzian signature, the null cone separates timelike directions from spacelike directions. The classification is invariant under any linear change of basis that preserves g (local Lorentz transformations), and therefore it is a geometric statement, not a coordinate dependent story.

Two algebraic observations are worth keeping in mind because they repeatedly appear later in disguised forms. First, the cone structure is stable under small perturbations of the metric: if you slightly change g, timelike vectors remain timelike and null vectors become either slightly timelike or slightly spacelike, but the basic separation into "inside", "on", and "outside" the cone persists. This stability is why one can talk about causality conditions that are robust under perturbations, such as stable causality [4]. Second, Lorentzian inner products satisfy a reversed Cauchy Schwarz type inequality for future directed timelike vectors u and v:

$$-g(u, v) \geq \sqrt{-g(u, u)}\,\sqrt{-g(v, v)}. \tag{3.2}$$

Geometrically, this inequality underlies the fact that timelike geodesics locally extremize (in fact, maximize) proper time. When we later discuss horizons and the behavior of

observers approaching them, the proper time functional is the quantity that keeps the physics honest while coordinates try to mislead.

3.2.2 Conceptual visual: the local light cone

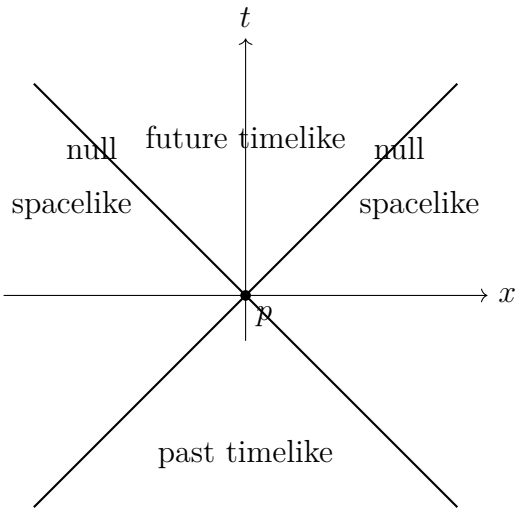


Figure 3.1: Conceptual light cone at an event p in a locally inertial chart. Timelike directions lie inside the cone, null directions lie on the cone, and spacelike directions lie outside.

3.2.3 Local inertial structure and why it matters

A Lorentzian metric does two jobs at once. It measures infinitesimal intervals and it sets the causal speed limit. The infinitesimal interval along a curve $\gamma(\lambda)$ is

$$\mathrm{d}s^2 = g(\dot\gamma, \dot\gamma)\,\mathrm{d}\lambda^2, \tag{3.3}$$

and a physically realizable worldline of a massive particle is modeled as a future directed timelike curve, meaning $g(\dot\gamma, \dot\gamma) < 0$ everywhere and $\dot\gamma$ stays in the chosen future cone. Proper time is the geometric length functional

$$\tau[\gamma] = \int_\gamma \sqrt{-g(\dot\gamma, \dot\gamma)}\,\mathrm{d}\lambda, \tag{3.4}$$

which is invariant under reparameterization. This invariance is the coordinate free statement that clocks do not care how you label the points on their worldlines.

The deepest reason Lorentzian geometry is tractable is that, at any point p, one can always choose local coordinates so that g equals the Minkowski metric at p and its first derivatives vanish at p. This is the local inertial property. Curvature reappears at second

order, and tidal effects are the geometry that survives coordinate erasure. But causality is already present at first order through the signature itself.

This separation between first order and second order information is a practical guide. If a statement is about the existence of light cones, the classification of vectors, or the possibility of choosing a future direction, it belongs to the signature level and can be analyzed without computing curvature. If a statement is about tidal stretching, focusing of geodesics, or the formation of conjugate points, it belongs to curvature and connection. The structure of the book follows this logic: we first build the causal and kinematic tools, and only then turn them into theorem machines via curvature identities such as Raychaudhuri's equation and the Einstein equations.

3.3 Light cones as local causal structure

3.3.1 Null curves and signal propagation

A curve γ is called **causal** if its tangent vector is never spacelike, meaning $g(\dot{\gamma}, \dot{\gamma}) \leq 0$ everywhere. Timelike curves model particles and clocks. Null curves model idealized light rays and, in general relativity, they also model gravitational wavefronts in the geometric optics limit. The phrase "information cannot travel faster than light" becomes the statement that physically meaningful influences propagate along causal curves.

The light cone at p is the set of null tangent vectors at p. It is a purely local object, but it induces a partial order on events: the causal future $J^+(p)$ consists of all points q that can be reached from p by a future directed causal curve. The chronological future $I^+(p)$ is defined similarly using timelike curves only. The distinction between $J^+(p)$ and $I^+(p)$ is not pedantry; it is the difference between what is reachable with a speed strictly less than light and what is reachable at or below light speed. Later, event horizons will be defined using J^+ and its behavior at infinity, which is why this chapter insists on the careful set theoretic language.

3.3.2 Conceptual visual: three vectors, three causal types

3.3.3 Why local cones do not settle global questions

Even though the light cone structure is defined pointwise, global causality is a global question. A spacetime can look perfectly well behaved in every local chart and still fail to support a sensible global time. For example, a spacetime may admit closed timelike curves, meaning there exists a smooth timelike curve that returns to its starting point.

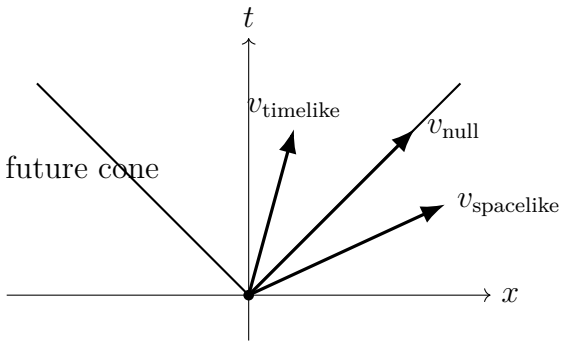


Figure 3.2: Conceptual classification of tangent vectors at a point. The null cone is the boundary separating timelike directions from spacelike directions.

No local measurement of the metric near any point detects that global pathology, because locally the spacetime is still Lorentzian with normal cones. This is why the subject has a hierarchy of causality conditions, from chronology (no closed timelike curves) to strong causality, to global hyperbolicity [4].

A concrete mental model is obtained by taking flat Minkowski space and identifying the time coordinate periodically, $t \sim t + T$. The local light cones at each point are unchanged, but the global identification turns the integral curves of ∂_t into closed timelike curves. This example is intentionally artificial, but it teaches a lesson that transfers to more subtle spacetimes: global causality is a statement about the manifold and the metric together, not about local formulas for g alone. In black hole physics, this is precisely why event horizons are defined using global causal relations with infinity rather than by inspecting a coordinate singularity at a particular radius.

3.4 Time orientation and the causal ladder

3.4.1 Time orientability

At each point p, timelike vectors form two disjoint cones. To speak about "future" and "past" consistently across $\mathcal{M}$, we need a global choice of one cone at every point that varies continuously. A Lorentzian manifold is **time orientable** if there exists a continuous timelike vector field T on $\mathcal{M}$. Such a field picks out one connected component of timelike vectors at each point, declared future directed by the requirement $g(T, v) < 0$ (with sign conventions adjusted as needed). Time orientability is a topological condition: it can fail even when the metric is smooth and nondegenerate everywhere.

Once time orientability is secured, one often wants more: a *time function* $t : \mathcal{M} \to \mathbb{R}$ that strictly increases along every future directed causal curve. Such a function provides

a global label that behaves like "time" even when no preferred Killing symmetry exists. Stable causality is equivalent to the existence of a time function with timelike gradient, and global hyperbolicity implies stable causality as part of the causal hierarchy [4]. In later chapters, when we draw Penrose diagrams, the ability to choose a global time function is what allows the diagram to encode causal relations without ambiguity.

Phenomenon statement: why time orientation is not optional

In astrophysical data analysis, every inference assumes that signals propagate from source to detector, not the reverse. That assumption is a time orientation. In the gravitational wave catalog used later in this chapter, each event is assigned a GPS time at the detector and an inferred redshift at the source. Those numbers are interpreted as lying on the detector's past light cone. Without a consistent time orientation, "past light cone" would not be globally meaningful, and even the idea of an event occurring before its detection would be coordinate dependent rather than geometric.

3.4.2 Causal hierarchy in one paragraph

The causal hierarchy is a nested family of conditions that forbid progressively more subtle causal pathologies. Chronology forbids closed timelike curves. Causality forbids closed causal curves (including null ones). Strong causality requires that small neighborhoods do not allow almost closed causal curves. Stable causality requires the existence of a time function whose gradient is everywhere timelike, ensuring that causality is robust under small perturbations of the metric. Global hyperbolicity is stronger still: it demands no causal pathologies and enough compactness of causal diamonds to guarantee well posed evolution [4]. Later chapters will treat event horizons as global objects, and global objects inevitably drag the causal hierarchy into the discussion.

3.5 Null hypersurfaces and degeneracy of induced metrics

3.5.1 What changes when a hypersurface is null

Let $\Sigma \subset \mathcal{M}$ be a smooth hypersurface, locally given as the level set $\Phi(x) = 0$ with $d\Phi \neq 0$ on Σ. The normal covector is $n_a = \nabla_a \Phi$. In Riemannian geometry, the induced metric on Σ is obtained by restricting g to tangent vectors of Σ, and it is always nondegenerate. In Lorentzian geometry this fails if Σ is **null**. A hypersurface is null if its normal covector is null with respect to g^{-1}, equivalently if the normal vector is tangent to the surface. In

that case, the restriction of g to $T\Sigma$ has a kernel: there exists a nonzero tangent vector k such that $g(k, w) = 0$ for all $w \in T\Sigma$. This is the geometric origin of the fact that a null hypersurface is "ruled" by null generators and cannot carry a standard Riemannian geometry intrinsically.

This degeneracy is not a defect. It is the feature that makes horizons special. Event horizons are null hypersurfaces generated by null geodesics. When we later talk about horizon area increase, expansion, shear, and the behavior of null congruences, we are exploiting this degenerate geometry rather than fighting it.

For readers who want a preview without jumping chapters: the geometry of a null hypersurface is often described by choosing a null generator field k tangent to the surface and an auxiliary transverse null vector ℓ with $g(k, \ell) = -1$. The induced metric on the two dimensional "screen space" orthogonal to both k and ℓ is nondegenerate and Riemannian. From it one defines the null second fundamental form and its trace, the expansion θ, which measures how cross sectional areas evolve along the generators. In a horizon context, θ is the quantity that appears in Raychaudhuri's equation and ultimately in the area theorem. The point here is structural: Lorentzian degeneracy is not an obstacle, it is the reason the horizon calculus has the form it does.

3.5.2 Conceptual visual: a null hypersurface with generators

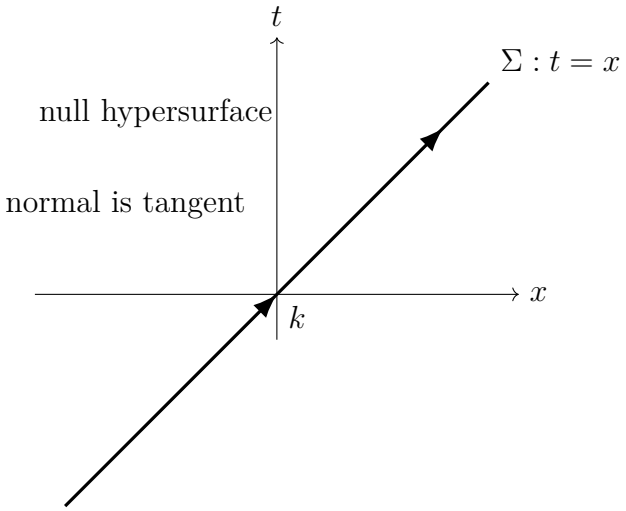


Figure 3.3: Conceptual picture of a null hypersurface in $1 + 1$ dimensions. The null direction k lies in the surface and is orthogonal (with respect to g) to every tangent direction in the surface, illustrating degeneracy of the induced metric.

3.6 Causal curves, domains of dependence, and global hyperbolicity

3.6.1 Causal futures and causal diamonds

Given a subset $S \subset \mathcal{M}$, define its causal future

$$J^+(S) = \bigcup_{p \in S} J^+(p), \tag{3.5}$$

and similarly its causal past $J^-(S)$. The **causal diamond** between two points p and q is $J^+(p) \cap J^-(q)$. In Minkowski space, causal diamonds are compact in the standard topology.

3.6.2 Conceptual visual: a causal diamond

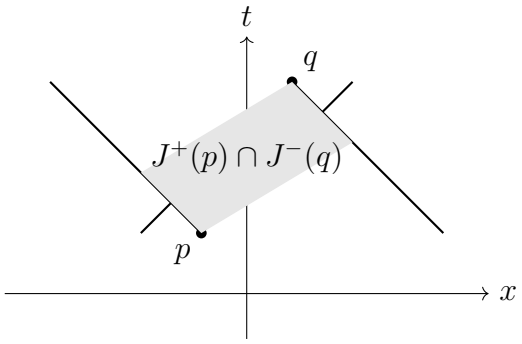


Figure 3.4: Conceptual causal diamond between p and q in $1+1$ dimensions, formed by intersecting the causal future of p with the causal past of q. Global hyperbolicity requires such diamonds to be compact for all p, q.

In curved spacetimes, that compactness can fail due to singularities or due to "holes" in the manifold. The compactness of causal diamonds is one of the two pillars of global hyperbolicity.

3.6.3 Domains of dependence

Let S be an achronal subset (no two points of S are timelike related). The **future domain of dependence** $D^+(S)$ is the set of points p such that every past inextendible causal curve through p intersects S. Intuitively, $D^+(S)$ is the region where initial data on S is sufficient to predict the future, because no signal can reach p without passing through

S. The past domain $D^-(S)$ is defined similarly, and the full domain of dependence is $D(S) = D^+(S) \cup D^-(S)$.

3.6.4 Conceptual visual: the domain of dependence

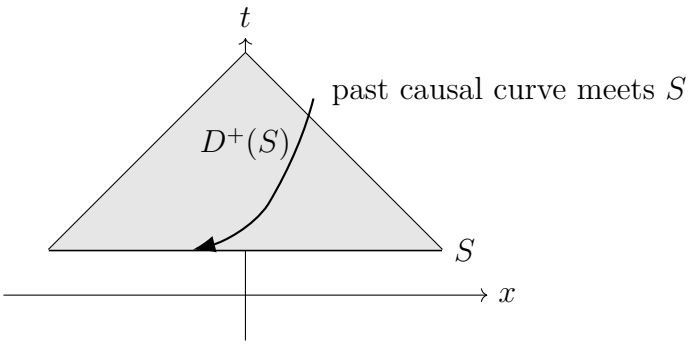


Figure 3.5: Conceptual domain of dependence. The future domain $D^+(S)$ consists of events whose entire past causal influence intersects the initial set S. The boundary is generated by null curves.

3.6.5 Global hyperbolicity and Cauchy surfaces

A spacetime $(\mathcal{M}, g)$ is **globally hyperbolic** if it is causal and if for every pair of points $p, q \in \mathcal{M}$ the set $J^+(p) \cap J^-(q)$ is compact. Equivalent characterizations exist. One of the most useful is the existence of a **Cauchy surface**, an achronal hypersurface S that every inextendible timelike curve intersects exactly once. When such a surface exists, the spacetime can be foliated by Cauchy surfaces and the Einstein equations admit a well posed initial value formulation on S (details in later chapters). The modern mathematical formulation of the foliation and splitting properties of globally hyperbolic spacetimes is developed in the work of Bernal and Sanchez [5, 6].

One particularly useful consequence is that a globally hyperbolic spacetime admits a smooth Cauchy time function t whose level sets are smooth Cauchy hypersurfaces. In many treatments this result is summarized as a "splitting": under mild assumptions $\mathcal{M}$ is diffeomorphic to $\mathbb{R} \times S$ with metric of the schematic form $-\alpha^2 \mathrm{d}t^2 + h_t$, where h_t is a Riemannian metric on the slice S that may vary with t. The important part is not the formula, but the meaning: global hyperbolicity provides a globally consistent notion of "initial data" and "evolution" without needing a Killing symmetry or asymptotic flatness.

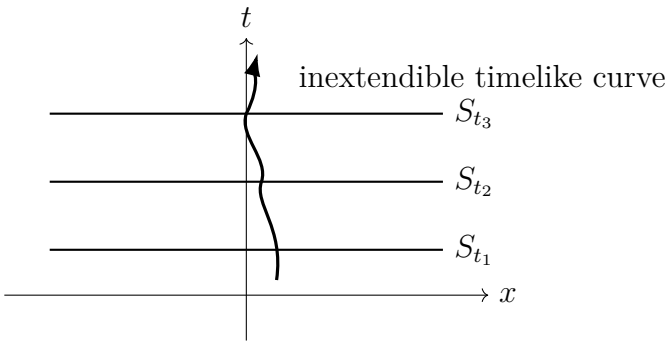


Figure 3.6: Conceptual foliation by Cauchy surfaces S_{t_i} in a globally hyperbolic spacetime. Every inextendible timelike curve intersects each slice exactly once, supporting deterministic evolution from initial data.

3.6.6 Conceptual visual: foliation by Cauchy slices

3.7 Data interlude: an empirical slice of the past light cone

3.7.1 Why a gravitational wave catalog belongs in a causality chapter

A gravitational wave detection is a causal story written in strain data and parameter estimation. The merger happens in a strong field region. The emitted gravitational radiation propagates outward along null directions. Eventually it reaches Earth, producing a correlated signal in detectors separated by thousands of kilometers. Even when we reduce the data to a catalog entry, that entry is still about null propagation: luminosity distance and redshift describe where the source sits on our past light cone, and source frame masses describe what happened in the strong field region.

To keep the figures transparent and reproducible in a static manuscript, we use a small subset of events extracted directly from the GWTC-4.0 default parameter CSV endpoint at GWOSC. The subset is included in this LaTeX file as a local table in `gwtc4_subset.csv`. Every data based plot in this chapter reads from that table, or from a histogram table that we computed by binning those same values [1].

Dataset note

The GWTC-4.0 catalog is continuously hosted and documented by the Gravitational Wave Open Science Center (GWOSC). The DOI for the GWTC-4.0 data release is included in each row of the CSV endpoint used here [2, 3]. This chapter uses only open values from that endpoint and simple derived quantities such as mass ratio $q = m_2/m_1$ and Schwarzschild radius $r_s \approx 2.95325\,\mathrm{km} \times (M/M_\odot)$.

3.7.2 Redshift and luminosity distance

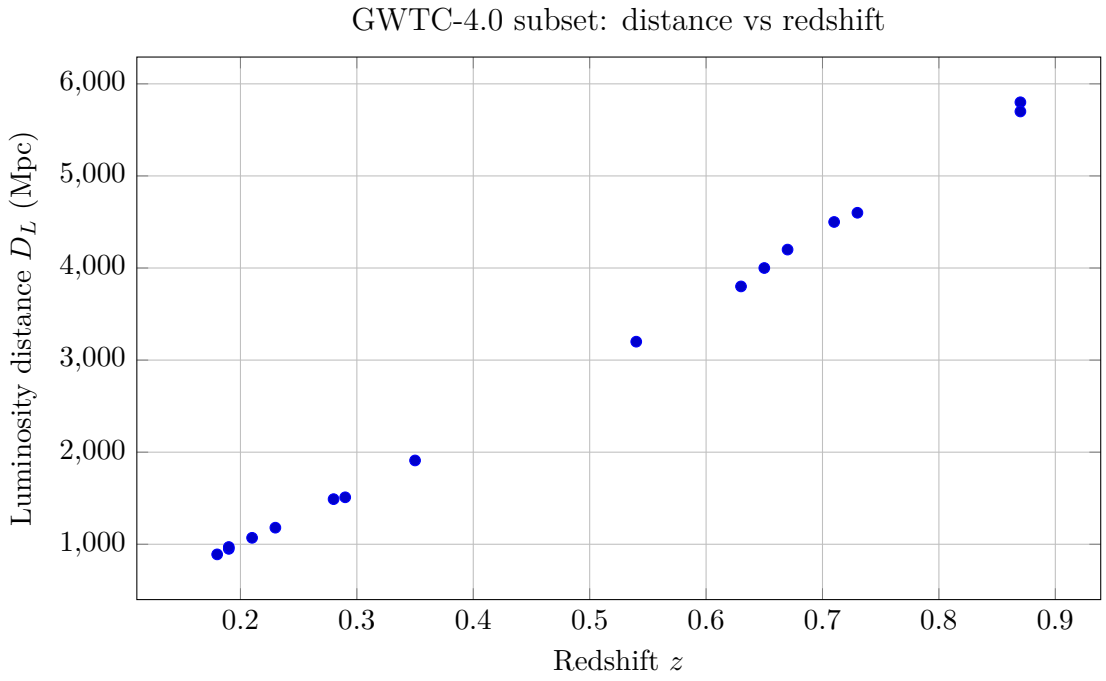


Figure 3.7: Observed event locations on the past light cone as summarized by catalog redshift and luminosity distance (subset). Each point is an inferred source location for a gravitational wave transient in the GWTC-4.0 public catalog [1].

The plot in Fig. 3.7 is not a cosmology tutorial. Its purpose is more geometric: it reminds us that every catalog event is a statement about null propagation from a distant spacetime region to the detector. In a Lorentzian manifold, a source at redshift z is not "far away" in a Euclidean sense. It is far along a null relation that connects emission to detection. Distance measures, redshift, and observational time all live on the light cone, and therefore they are best interpreted with causal geometry in mind.

3.7.3 Chirp mass as a geometry sensitive parameter

The chirp mass M_c is the combination of component masses that controls the leading order phase evolution of the inspiral. It is defined by

$$M_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}. \tag{3.6}$$

While this chapter does not develop waveform theory, the appearance of M_c in a causality chapter is appropriate: detectors measure a time series, and the strongest part of the inference comes from the way phase accumulates as a function of time. Time is a Lorentzian concept, and the entire inference pipeline is a practical demonstration that the geometry of causal curves is not optional if you want to interpret data from strong field regions.

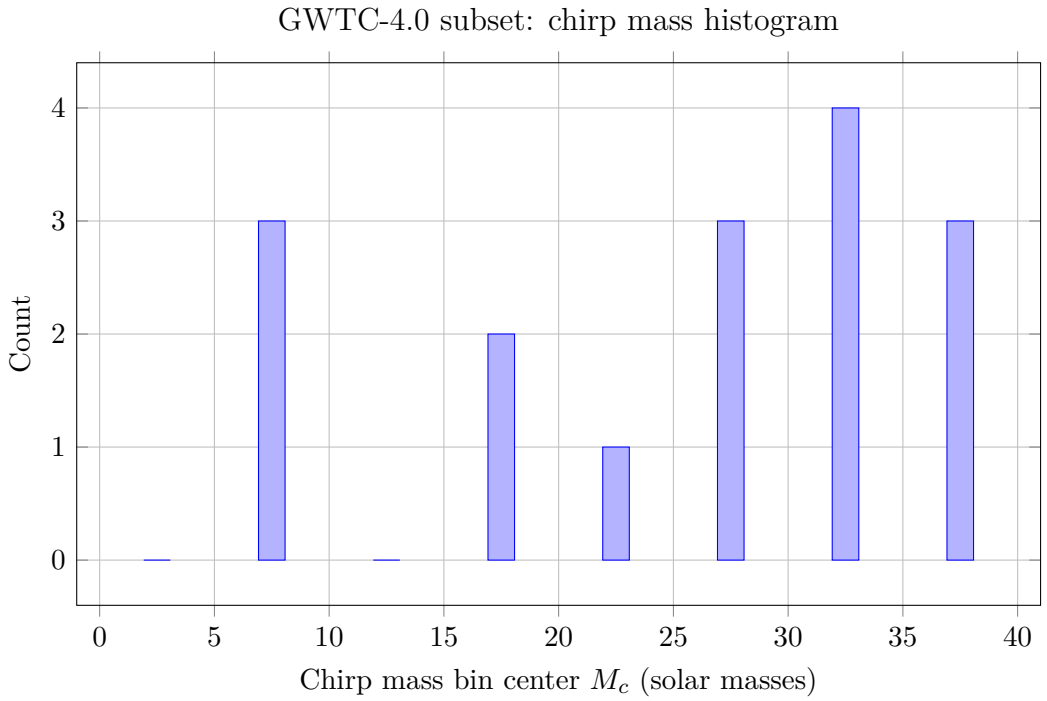


Figure 3.8: Histogram of chirp mass values M_c in the embedded GWTC-4.0 subset, binned in 5 solar mass intervals. Values originate from the GWTC-4.0 default parameter endpoint [1].

3.7.4 Effective spin and the geometry of precession

The effective spin parameter χ_{eff} is a particular projection of the component spins onto the orbital angular momentum. It is bounded between -1 and 1 for Kerr black holes in general relativity and is especially relevant because it is relatively well measured compared with individual spin components. Although spin dynamics will be revisited in the Kerr chapters, it is worth highlighting here that spin enters gravitational wave phasing through

couplings that depend on the causal structure of the orbit. Precession is a geometry of frames along timelike curves, and timelike curves are Lorentzian objects.

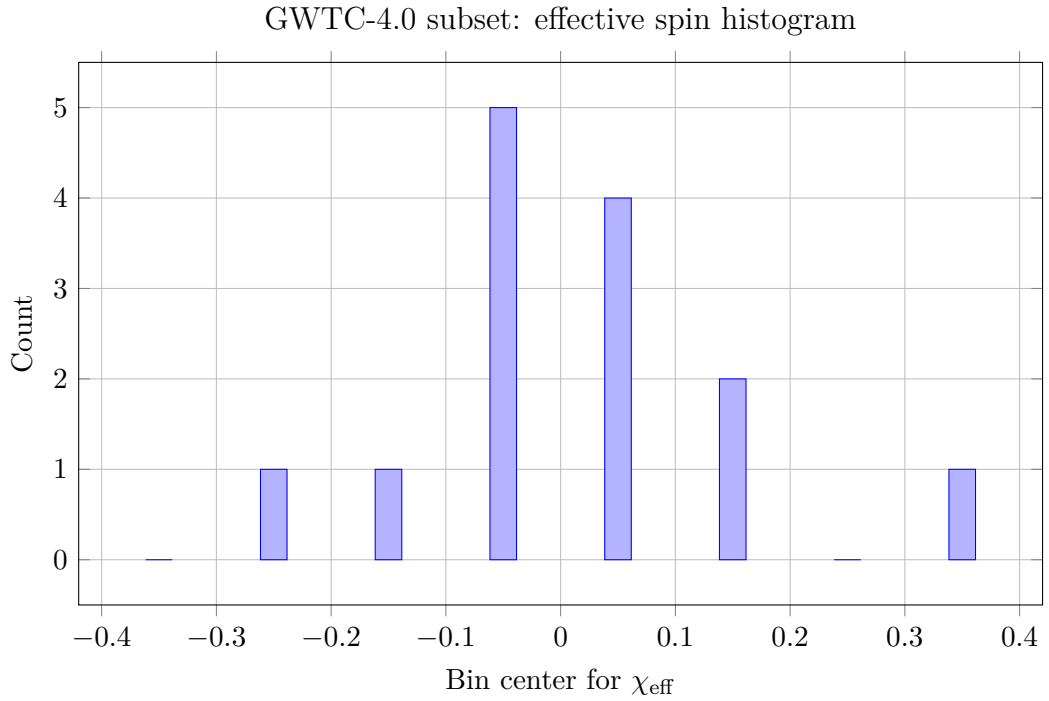


Figure 3.9: Histogram of effective spin χ_{eff} values in the embedded GWTC-4.0 subset, binned in steps of 0.1. Values originate from the GWTC-4.0 default parameter endpoint [1].

3.7.5 Component masses and mass ratio

3.7.6 Detectability: signal to noise ratio versus distance

3.7.7 From total mass to a horizon scale

Even though detailed horizon geometry is postponed to later chapters, the catalog parameters already suggest a natural length scale. For a nonrotating black hole of mass M, the Schwarzschild radius is

$$r_s = \frac{2GM}{c^2} \approx 2.95325 \text{ km} \times \frac{M}{M_\odot}. \tag{3.7}$$

Using the total mass inferred for each event, we can compute an approximate horizon scale for the system. This conversion is purely algebraic. It is not a simulation. It is the same kind of dimensional translation that turns a measured wavelength into a frequency.

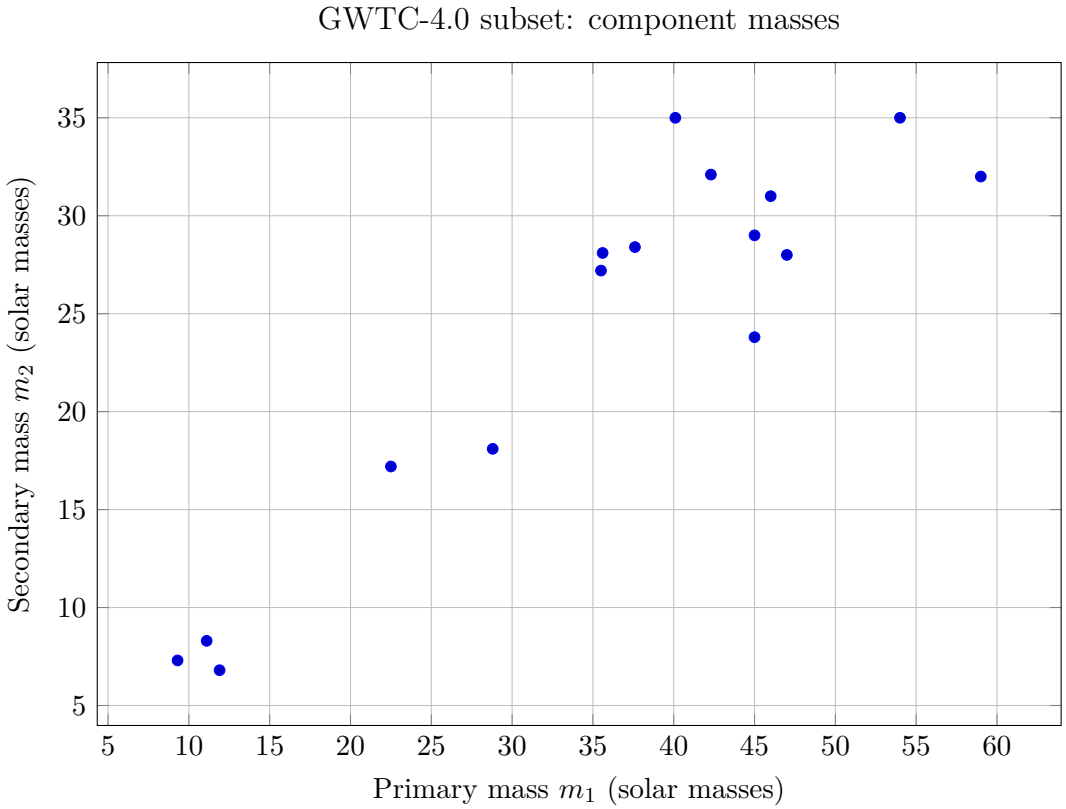


Figure 3.10: Component masses for events in the embedded GWTC-4.0 subset. These masses are inferred in the source frame and summarize the strong field system that produced the observed null signal [1].

3.7.8 Final mass and radiated energy proxy

Gravitational waves carry energy away, so the final remnant mass is smaller than the initial total mass. The difference between the inferred total mass and inferred final mass is an observational proxy for radiated energy, up to the conversion factor c^2 and inference uncertainties. From a Lorentzian viewpoint, the important lesson is that energy is not just a number: it is a component of the stress energy tensor, and it is constrained by causal propagation and conservation laws.

3.7.9 Statistical significance: false alarm rate distribution

The causal story of a detection ends not only in parameter inference but also in significance estimation. The false alarm rate (FAR) is an operational proxy for how often noise alone would produce a candidate of similar ranking statistic. It is not pure geometry, but it is a reminder that causal inference from data requires a robust definition of what counts as a real signal.

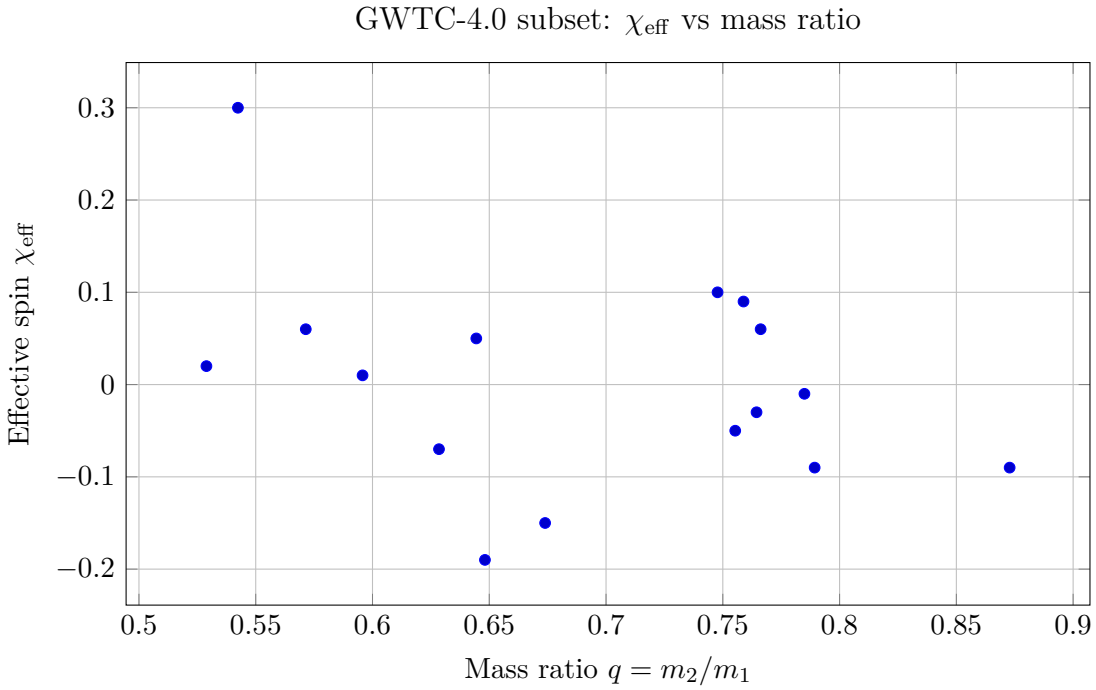


Figure 3.11: Effective spin versus mass ratio for the embedded GWTC-4.0 subset. Both quantities are inferred from causal phase evolution in detector time series data [1].

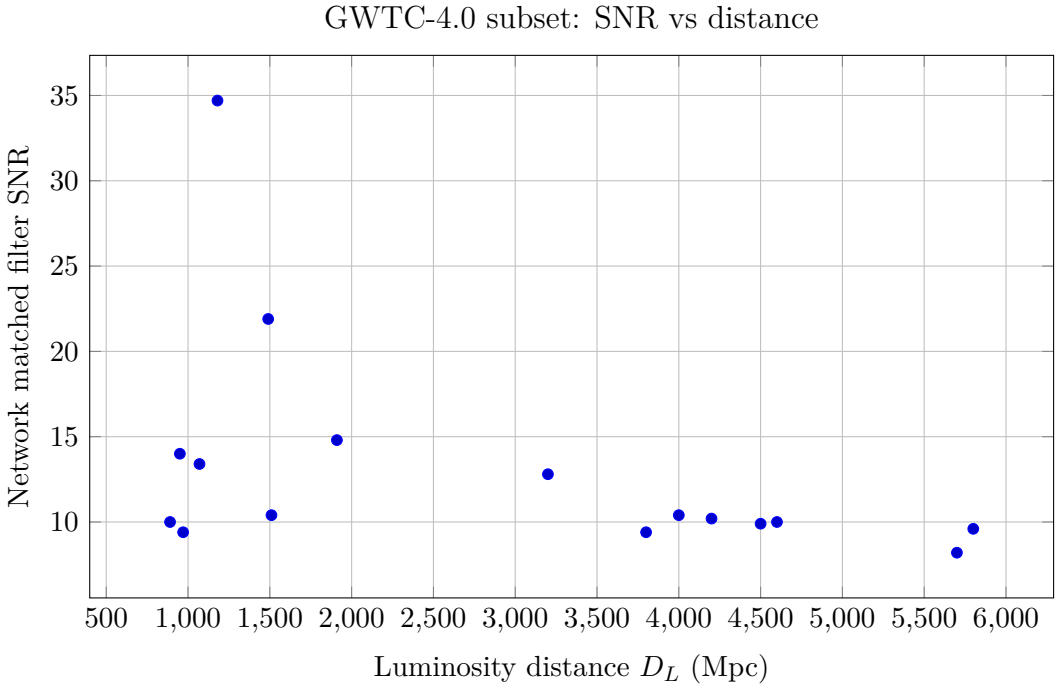


Figure 3.12: A detectability snapshot: network matched filter signal to noise ratio (SNR) versus luminosity distance for the embedded GWTC-4.0 subset. This is a practical reminder that what reaches us along null curves arrives with amplitude that decays with distance and is filtered through detector noise [1].

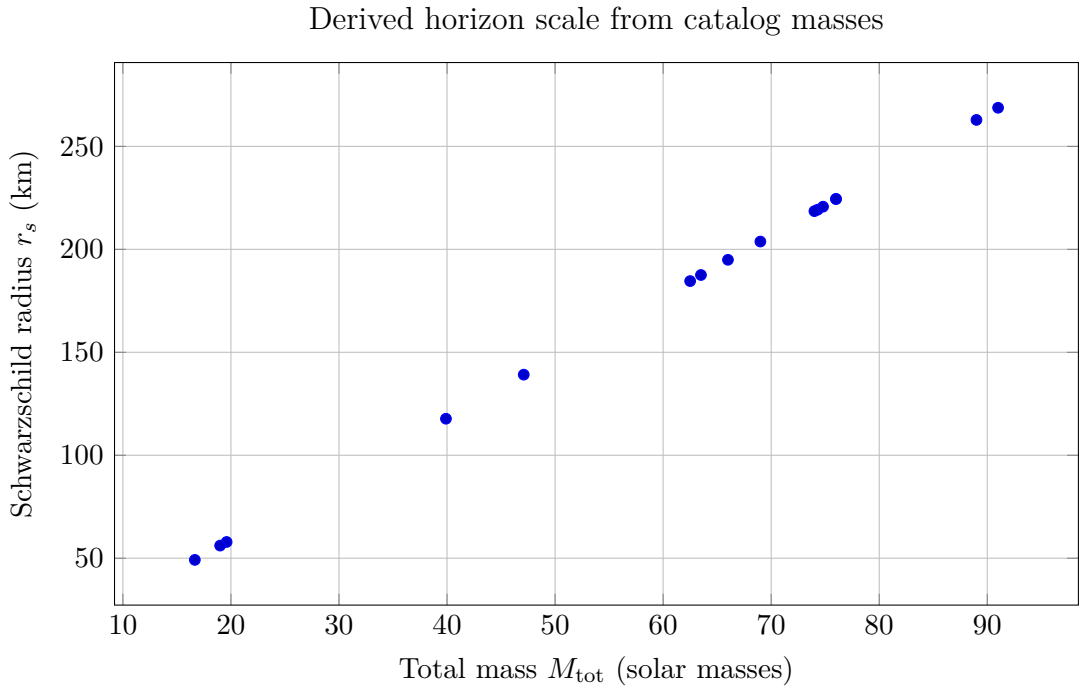


Figure 3.13: A derived horizon scale computed from catalog total mass using $r_s \approx 2.95325\,\mathrm{km} \times (M/M_\odot)$. Masses come from the embedded GWTC-4.0 subset [1].

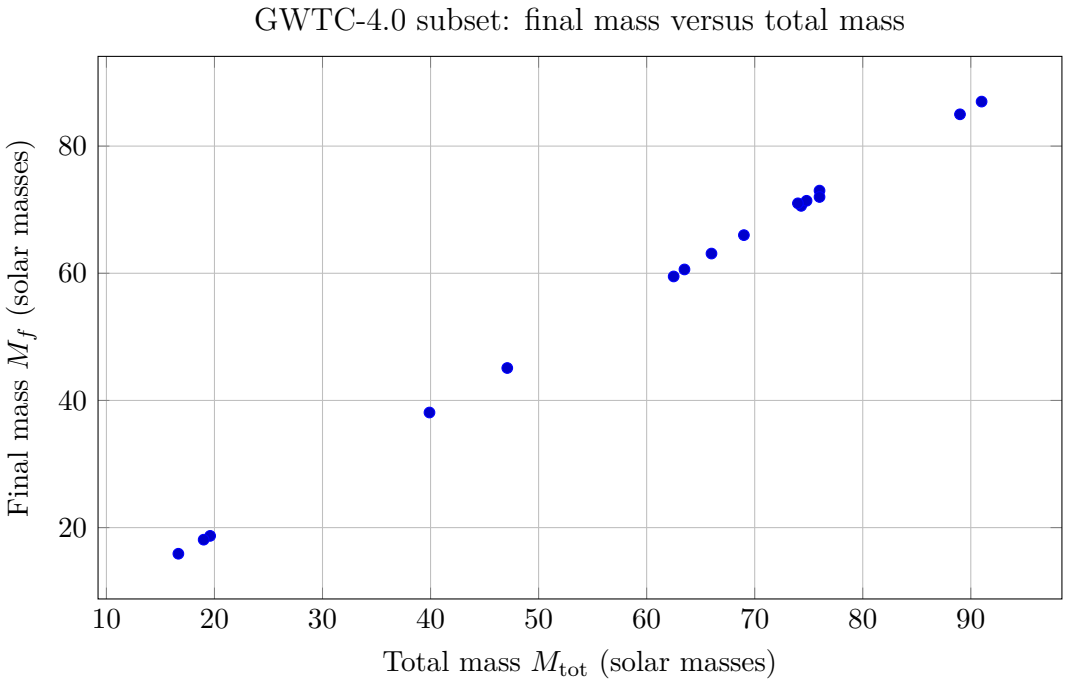


Figure 3.14: Final mass versus total mass for the embedded GWTC-4.0 subset, illustrating that a fraction of the system mass energy is radiated away as gravitational waves that propagate along null curves [1].

3.7.10 A minimal time ordered view

Catalog entries are also time stamped at the detector. The GPS time is a coordinate time used for synchronization of detectors on Earth. While GPS time is not a geometric

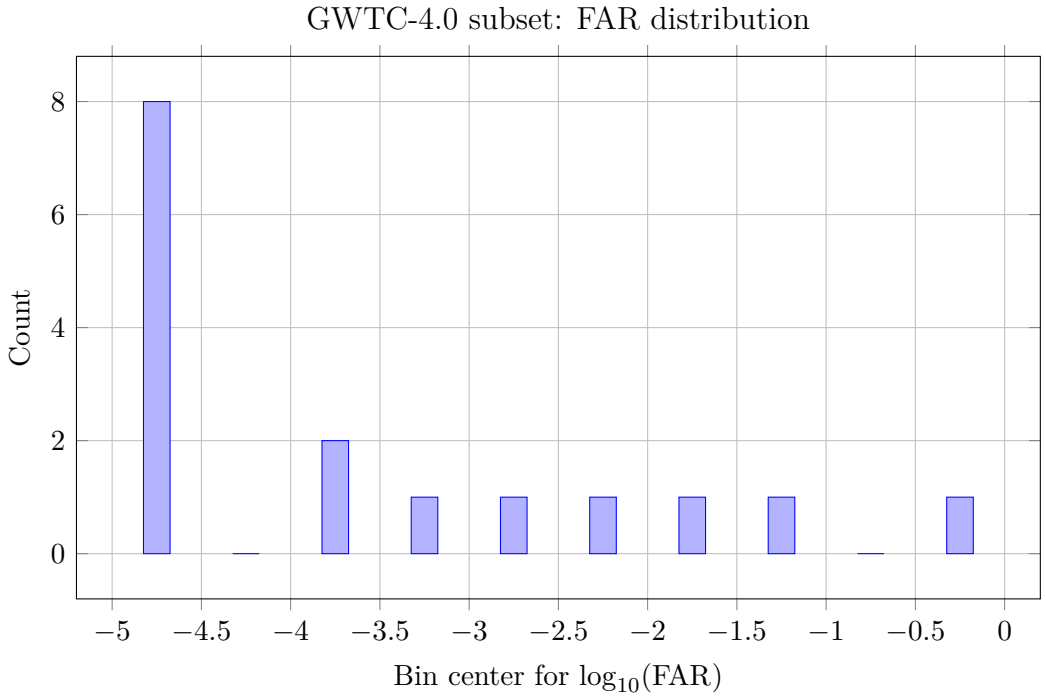


Figure 3.15: Histogram of $\log_{10}$(FAR) values for the embedded GWTC-4.0 subset. FAR values come directly from the GWTC-4.0 default parameter endpoint [1].

invariant, it is an operationally consistent parameter in the data pipeline. Sorting events by GPS time and plotting the cumulative count provides a simple empirical picture of how observations accumulate over detector time.

3.8 Real world applicability: phenomenon statements

The definitions above are abstract by design. They are written so they still mean the same thing when we switch coordinate systems, when we move from Minkowski space to Schwarzschild to Kerr, and when we talk about global questions like horizons. But the same definitions also show up in the laboratory and in the sky.

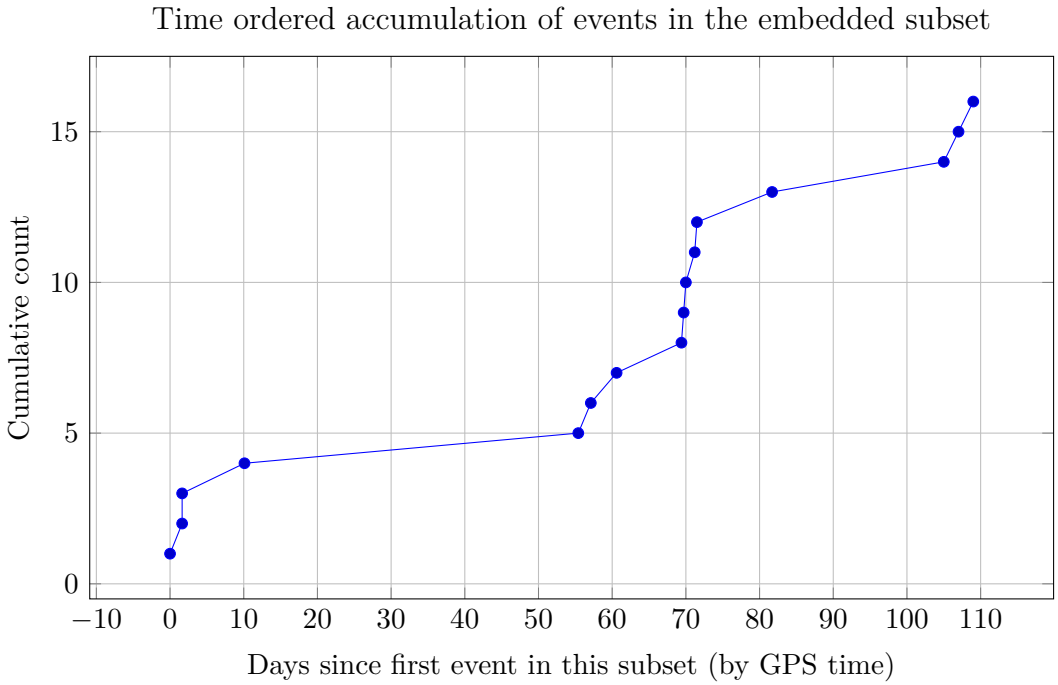


Figure 3.16: Cumulative number of events versus days since the earliest GPS time in the embedded GWTC-4.0 subset. Times are taken from the catalog and differences are computed algebraically [1].

Phenomenon statement: GPS as Lorentzian geometry in infrastructure

A GPS satellite clock follows a timelike worldline in Earth's spacetime. The receiver clock on the ground follows a different timelike worldline. If you model Earth as a static gravitational field and the satellite as moving relative to the ground, both gravitational time dilation and special relativistic time dilation appear. The engineering pipeline is a practical computation of proper time along two timelike curves and the translation between proper time and a coordinate time used for synchronization. The conceptual content is the same as in this chapter: timelike curves carry clocks, null curves carry signals, and the partial order induced by the light cones determines which events can be compared by exchanged signals.

Phenomenon statement: lensed quasar time delays as null geodesic bookkeeping

In strong gravitational lensing, multiple images of the same variable source appear because light follows multiple null geodesics around a lensing mass. Different paths accumulate different geometric lengths and different gravitational time delays. The measured arrival time difference between image light curves is literally a difference between two null related causal histories. Even before we learn the full optical geometry machinery, Lorentzian causality already predicts what must be true: all measured images are connected to the same emission event by null curves, and the relative delay is a property of the null congruence structure of the spacetime.

Phenomenon statement: gravitational wave networks as causal triangulators

Two detectors separated by a baseline cannot see the same transient at exactly the same detector time unless the source is located on the perpendicular bisector of the baseline with respect to the relevant wavefront. A measured arrival time delay is therefore a direct constraint on the direction of the incoming null wavefront. This is light cone geometry turned into an instrument. The catalog parameters plotted earlier are the final product of this causal triangulation plus waveform inference. The geometry is not decoration. It is the thing that makes the inference possible.

3.9 Conclusion

Lorentzian geometry is the grammar of relativistic physics. The signature of the metric is a local statement, but it creates global consequences: a causal structure, a notion of time orientation, and the possibility (or failure) of deterministic evolution from initial data. Null hypersurfaces are not exotic objects introduced only for black holes; they arise naturally whenever a hypersurface is generated by lightlike directions, and their induced metric degeneracy is a feature that will become central when horizons enter the discussion.

The chapter also emphasized that these concepts are not sealed inside pure mathematics. Public datasets from gravitational wave astronomy encode causal structure in a practical form: distances and redshifts locate sources on the past light cone, and the temporal phasing of signals carries information about source masses and spins, which in turn foreshadow the horizon scales studied later in this book. When later chapters speak about event horizons as global boundaries defined through causal futures and null infinity, the definitions will not be new. They will be the same definitions introduced here, now applied to spacetimes where the geometry is strong enough to hide regions from the rest of the universe.

Bibliography

[1] Gravitational Wave Open Science Center (GWOSC), "GWTC-4.0 Event Catalog, default parameters (CSV endpoint)," accessed 2026-02-24. [Online]. Available: https://gwosc.org/api/v2/catalogs/GWTC-4.0/events?format=csv&include-default-parameters=true

[2] Gravitational Wave Open Science Center (GWOSC), "GWTC-4.0 Data Release DOI: 10.7935/aes8-px89," [Online]. Available: https://doi.org/10.7935/aes8-px89

[3] Gravitational Wave Open Science Center (GWOSC), "GWOSC API documentation," [Online]. Available: https://gwosc.org/api/

[4] E. Minguzzi and M. Sanchez, "The causal hierarchy of spacetimes," Living Reviews in Relativity, vol. 11, no. 2, 2008. doi: 10.12942/lrr-2008-2. [Online]. Available: https://doi.org/10.12942/lrr-2008-2

[5] A. N. Bernal and M. Sanchez, "Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes," Communications in Mathematical Physics, 2005. [Online]. Available: https://arxiv.org/abs/gr-qc/0512095

[6] A. N. Bernal and M. Sanchez, "On smooth Cauchy hypersurfaces and Geroch's splitting theorem," 2003. [Online]. Available: https://arxiv.org/abs/gr-qc/0306108

[7] B. P. Abbott *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger," Physical Review Letters, vol. 116, no. 6, p. 061102, 2016. doi: 10.1103/PhysRevLett.116.061102. [Online]. Available: https://doi.org/10.1103/PhysRevLett.116.061102

[8] R. M. Wald, *General Relativity*. Chicago, IL, USA: University of Chicago Press, 1984.

[9] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*. Cambridge, UK: Cambridge University Press, 1973. [Online]. Available: https://www.amazon.com/Structure-Space-Time-Cambridge-Monographs-Mathematical/dp/0521200164

Contents

4 Symmetries as Geometry: Killing Fields and Conserved Quantities **1**

Abstract . . . 1
Keywords . . . 1
4.1 Introduction . . . 1
4.2 Isometries and Killing vector fields . . . 2
4.2.1 Isometries as metric-preserving diffeomorphisms . . . 2
4.2.2 One-parameter isometry groups and the Killing equation . . . 3
4.2.3 What Killing fields buy you . . . 3
4.3 Lie derivatives of tensors . . . 4
4.3.1 Definition via the flow . . . 4
4.3.2 The Killing equation revisited . . . 4
4.4 Noether's theorem in geometric form . . . 5
4.4.1 From invariance to conservation . . . 5
4.4.2 Noether currents for fields and geometric caveats . . . 6
4.5 Symmetry-adapted frames and coordinates . . . 6
4.6 Hidden symmetries: Killing tensors and the Carter constant . . . 6
4.7 Data-driven illustrations from an open catalog . . . 8
4.7.1 Dataset provenance and derived quantities . . . 8
4.7.2 Index table of the variables used in plots . . . 9
4.8 Ten data-based illustrations . . . 9
4.9 Additional conceptual visuals to support intuition . . . 15
4.10 Real-world applicability: phenomenon statements . . . 15
4.10.1 Phenomenon 1: gravitational wave chirps as evolving symmetry charges . . . 16
4.10.2 Phenomenon 2: black hole shadows as symmetry-filtered null geodesics . . . 16
4.10.3 Phenomenon 3: stellar orbits around Sgr A* as a laboratory for integrals of motion . . . 17
4.10.4 Phenomenon 4: accretion disk spectroscopy as a measurement of spacetime symmetries . . . 17
4.10.5 Phenomenon 5: relativistic jets as rotational energy extraction . . . 17
4.10.6 Phenomenon 6: ringdown as symmetry restoration . . . 17
4.11 Conclusion . . . 18

List of Figures

4.1 Conceptual visual (non-dataset): a one-parameter family of isometries generates flow lines (orbits) on the manifold $\mathcal{M}$. The tangent to these orbits is the Killing field ξ^a. 3

4.2 Conceptual visual (non-dataset): the Lie derivative measures the infinitesimal change of a tensor field under pullback by the flow generated by ξ^a. 4

4.3 Conceptual visual (non-dataset): Noether's theorem specialized to geodesic motion. A metric symmetry generated by a Killing field implies invariance of the geodesic action and yields a conserved scalar along each geodesic. 6

4.4 Conceptual visual (non-dataset): symmetry-adapted frames align basis vectors with Killing directions. In rotating spacetimes, the geometry forces mixing between the Killing directions and the transverse directions, and the mixed terms carry physical meaning. 7

4.5 Conceptual visual (non-dataset): visible symmetries generated by Killing vectors produce linear integrals of motion; hidden symmetries encoded by Killing tensors produce quadratic integrals. Kerr spacetime exhibits both. 8

4.6 Data illustration 1 (GWTC-3-confident): component masses inferred from gravitational waves, colored by effective spin $\chi_{\rm eff}$. The diagonal line marks equal-mass binaries. Data from GWOSC Event Portal CSV [2]. 10

4.7 Data illustration 2 (GWTC-3-confident): chirp mass $\mathcal{M}$ versus total mass $M_{\rm tot}$. The straight lines are theoretical relations $\mathcal{M} = M_{\rm tot}\,\eta^{3/5}$ for fixed mass ratios (approximate slopes shown for $q = 1$, $q \approx 0.25$, and $q \approx 0.10$). Data from GWOSC [2]. 10

4.8 Data illustration 3 (GWTC-3-confident): histogram of total source-frame mass $M_{\rm tot}$. Data from GWOSC [2]. 11

4.9 Data illustration 4 (GWTC-3-confident): histogram of mass ratio q. Data from GWOSC [2]. 11

4.10 Data illustration 5 (GWTC-3-confident): histogram of effective inspiral spin $\chi_{\rm eff}$. This parameter is a particular projection and mass-weighted combination of component spins inferred from waveform phasing. Data from GWOSC [2]. 12

4.11 Data illustration 6 (GWTC-3-confident): histogram of luminosity distance. Selection effects strongly shape the observed distribution. Data from GWOSC [2]. . . 13

4.12 Data illustration 7 (GWTC-3-confident): network matched-filter SNR versus luminosity distance in log-log scale. The curve is a best-fit power law $\rho \approx 67.583\,D_L^{-0.2496}$ computed from the plotted data (fit shown only as a descriptive trend; detector sensitivity and selection effects complicate the scaling). Data from GWOSC [2]. 13

4.13 Data illustration 8 (GWTC-3-confident): remnant mass versus initial total mass. The diagonal $M_{\mathrm{final}} = M_{\mathrm{tot}}$ corresponds to no radiated energy, while the dashed line is a least-squares linear fit $M_{\mathrm{final}} \approx 0.9520\, M_{\mathrm{tot}} + 0.2513$ for the plotted events. Data from GWOSC [2]. 14
4.14 Data illustration 9 (GWTC-3-confident): radiated energy fraction $f_{\mathrm{rad}} = \big((m_1 + m_2) - M_{\mathrm{final}}\big)/(m_1 + m_2)$ as a function of symmetric mass ratio η. The curve is a quadratic least-squares fit to the plotted data. Data from GWOSC [2]. 14
4.15 Data illustration 10 (GWTC-3-confident): effective spin versus mass ratio with asymmetric credible intervals (error bars) taken from the catalog offsets. The solid line is a least-squares linear fit $\chi_{\mathrm{eff}} \approx -0.1633\, q + 0.1516$ for the plotted medians. Data from GWOSC [2]. 15
4.16 Conceptual visual (non-dataset): conserved quantities carve phase space into lower-dimensional surfaces. More independent integrals mean motion is confined more strongly, and integrability becomes plausible. 16

Chapter 4

Symmetries as Geometry: Killing Fields and Conserved Quantities

Abstract

Symmetry is a way of saying that the metric contains redundancy. Some directions in spacetime are "free" in the precise sense that sliding the entire geometry along those directions changes nothing. When such an invariance exists, it is not merely an aesthetic property. It becomes a computational engine that turns a difficult differential equation into an algebraic constant of motion. In this chapter we treat symmetries as geometric objects, generated by Killing vector fields and, more subtly, by higher rank Killing tensors. We develop the Killing equation through Lie differentiation, connect invariance of an action to conserved currents via Noether's theorem, and explain why stationary and axisymmetric black hole spacetimes are solvable to a remarkable degree. To anchor the discussion in observation rather than pure formalism, we include ten data-based illustrations built from the open GWTC-3-confident gravitational wave catalog provided by the Gravitational Wave Open Science Center. From the catalog medians and credible intervals we compute derived quantities such as mass ratio, symmetric mass ratio, and radiated energy fraction, and use these to visualize how "symmetry-coded" charges like mass and angular momentum manifest in real astrophysical populations. The goal is to build geometric intuition that will later reappear as a practical tool in Schwarzschild and Kerr geometry.

Keywords

Killing vectors, isometries, Lie derivative, Noether theorem, conserved quantities, symmetry-adapted frames, Killing tensors, Carter constant, integrability, gravitational wave catalogs.

4.1 Introduction

Symmetry is sometimes introduced as a convenience: a shortcut that reduces the number of coordinates you have to keep track of. In geometry, symmetry is deeper than convenience. An isometry is not a coordinate choice; it is a statement about the manifold itself. It is the existence of a diffeomorphism that leaves the metric unchanged. If you picture a spacetime as a fabric, an isometry is a rigid motion of that fabric in which every ruler and every clock built from the fabric continues to agree with itself. This kind of invariance is rare in generic curved spacetimes, but

when it occurs it produces an immediate dividend: a conserved quantity along geodesics. This fact is the geometric face of Noether's theorem, and it is one reason black hole spacetimes are so mathematically tractable. The astrophysical black holes we observe are complicated systems interacting with messy matter, magnetic fields, and galaxies. Yet the region close to the horizon is often well approximated by an exact metric with high symmetry, and that symmetry can be read in the data as a small set of parameters.

The chapter title promises a shift in mindset. We do not treat symmetry as a property that happens after you write down a metric in a convenient chart. Instead we treat symmetry as a vector field or tensor field with an intrinsic definition. The main actor is a Killing vector field ξ^a, defined by the geometric condition $\mathcal{L}_\xi g_{ab} = 0$. This single equation encodes time translation invariance, rotational invariance, and more exotic invariances like boosts in Minkowski spacetime. The corresponding conserved quantity is $\xi_a u^a$ along a geodesic with tangent u^a. The same pattern repeats at higher rank: a Killing tensor K_{ab} obeys $\nabla_{(a}K_{bc)} = 0$, and produces a quadratic constant $K_{ab}u^a u^b$. In Kerr geometry this "hidden" symmetry yields the Carter constant, a fact that will later make the Kerr geodesic equations separable and its photon regions understandable.

Symmetry, then, is a bridge between differential geometry and physics: it translates between the Lie algebra of isometries and the algebra of constants of motion. This bridge is not abstract. It is the reason gravitational wave data analysis can compress the complicated two body problem into a handful of parameters, and it is the reason the end state of a binary black hole merger is described so accurately by two numbers, mass and spin, which are themselves Noether charges associated with Killing fields of the remnant Kerr spacetime. In the second half of this chapter we use the open GWTC-3-confident catalog of compact binary mergers to build data-based plots that illustrate these ideas in an observationally grounded way [1, 2].

4.2 Isometries and Killing vector fields

4.2.1 Isometries as metric-preserving diffeomorphisms

Definition 4.1 (Isometry)**.** *Let $(\mathcal{M}, g)$ be a (pseudo-)Riemannian manifold. A diffeomorphism $\varphi : \mathcal{M} \to \mathcal{M}$ is an* isometry *if*

$$\varphi^* g = g. \tag{4.1}$$

Equivalently, g_{ab} and its pullback by φ agree as tensor fields on $\mathcal{M}$.

This definition is coordinate-free. It says that all inner products of all pairs of vectors are preserved when you push them forward by φ. If v^a and w^a are tangent vectors at a point $p \in \mathcal{M}$, then

$$g_{ab}(p)\, v^a w^b = g_{ab}(\varphi(p))\, (\varphi_* v)^a (\varphi_* w)^b. \tag{4.2}$$

In Lorentzian geometry, this isometry condition respects not only lengths but also causal character. Timelike, null, and spacelike vectors remain timelike, null, and spacelike. In a physics interpretation, an isometry represents a transformation that does not change any local experiment performed with idealized rulers and clocks. A concise coordinate-free discussion of isometries, flows, and Killing fields can be found in standard general relativity lecture notes [8, 9].

Many spacetimes have no nontrivial isometries. A generic curved metric has no continuous symmetries at all. This is not a defect of the definition; it is a reflection of how much information a general metric contains. When a spacetime *does* have symmetries, the set of all isometries forms a Lie group, and the infinitesimal generators of that group are the Killing vector fields.

4.2.2 One-parameter isometry groups and the Killing equation

A particularly useful situation is when the isometries come in a smooth one-parameter family $\{\varphi_\lambda\}$ with $\varphi_0 = \mathrm{id}$ and $\varphi_{\lambda+\mu} = \varphi_\lambda \circ \varphi_\mu$. The derivative at $\lambda = 0$ defines a vector field:

$$\xi^a(p) := \left.\frac{\mathrm{d}}{\mathrm{d}\lambda}\right|_{\lambda=0} \varphi_\lambda(p). \tag{4.3}$$

The field ξ^a is the infinitesimal generator of the symmetry. The metric preservation condition $\varphi_\lambda^* g = g$ differentiates to

$$\left.\frac{\mathrm{d}}{\mathrm{d}\lambda}\right|_{\lambda=0} \varphi_\lambda^* g = \mathcal{L}_\xi g = 0. \tag{4.4}$$

Definition 4.2 (Killing vector field)**.** *A vector field ξ^a is a* Killing field *if*

$$\mathcal{L}_\xi g_{ab} = 0. \tag{4.5}$$

In terms of the Levi-Civita connection ∇,

$$\nabla_a \xi_b + \nabla_b \xi_a = 0. \tag{4.6}$$

Equation (4.6) is the Killing equation. It states that the covariant derivative of ξ_a is antisymmetric. Intuitively, the flow generated by ξ^a drags points along symmetry orbits, but it does so without stretching or shearing the metric. Another useful way to think is local: at each point, $\nabla_a \xi_b$ acts like an infinitesimal Lorentz transformation on tangent vectors.

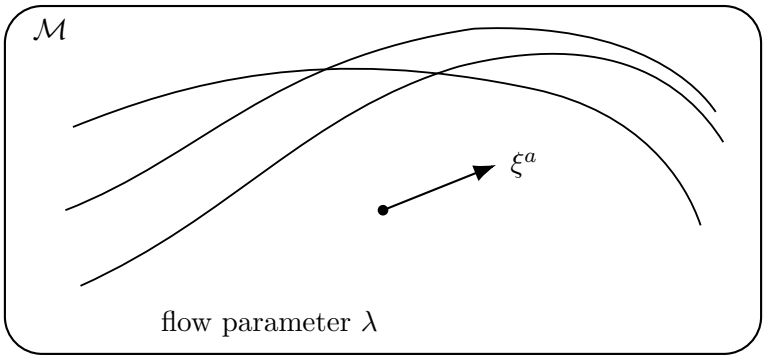


Figure 4.1: Conceptual visual (non-dataset): a one-parameter family of isometries generates flow lines (orbits) on the manifold $\mathcal{M}$. The tangent to these orbits is the Killing field ξ^a.

4.2.3 What Killing fields buy you

The Killing equation has immediate consequences. First, the set of Killing fields forms a Lie algebra under the Lie bracket: if ξ and η are Killing, then so is $[\xi, \eta]$. This is why symmetries organize into groups. Second, Killing fields constrain the metric coefficients. If you manage to find a coordinate chart adapted to ξ so that $\xi = \partial/\partial x^0$, then $\mathcal{L}_\xi g = 0$ implies that g_{ab} is independent of x^0. This is the coordinate expression of stationarity or axisymmetry.

Third, and most important for physics, Killing fields yield conserved quantities along geodesics, which is the content of the next section. But before reaching conservation laws, we need a solid geometric understanding of the Lie derivative, because the Killing equation is simply $\mathcal{L}_\xi g = 0$.

4.3 Lie derivatives of tensors

4.3.1 Definition via the flow

Let $\{\varphi_\lambda\}$ be the flow of a vector field ξ^a. For any tensor field T on $\mathcal{M}$, the Lie derivative along ξ is defined by

$$\mathcal{L}_\xi T := \left.\frac{\mathrm{d}}{\mathrm{d}\lambda}\right|_{\lambda=0} \varphi_\lambda^* T. \tag{4.7}$$

This definition has three features that make it the natural derivative for symmetry.

1. It is coordinate-free: it depends only on the flow.
2. It measures change under active transport of points by φ_λ.
3. It is the correct notion of derivative for comparing tensors at different points without introducing extra structure beyond the diffeomorphism itself.

When a connection is available, the Lie derivative can be written in local form. For a scalar f, $\mathcal{L}_\xi f = \xi^a \nabla_a f$. For a vector v^a,

$$\mathcal{L}_\xi v^a = \xi^b \nabla_b v^a - v^b \nabla_b \xi^a. \tag{4.8}$$

For a covector ω_a,

$$\mathcal{L}_\xi \omega_a = \xi^b \nabla_b \omega_a + \omega_b \nabla_a \xi^b. \tag{4.9}$$

A detailed derivation of these identities and their relation to flows and commutators is given, for example, in [8]. These formulas extend to general tensors by the Leibniz rule and by insisting that $\mathcal{L}_\xi$ commute with contraction.

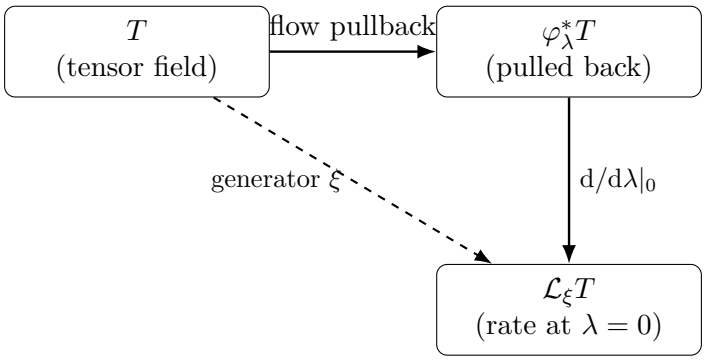


Figure 4.2: Conceptual visual (non-dataset): the Lie derivative measures the infinitesimal change of a tensor field under pullback by the flow generated by ξ^a.

4.3.2 The Killing equation revisited

Applying the vector formula to the metric yields

$$(\mathcal{L}_\xi g)_{ab} = \xi^c \nabla_c g_{ab} + g_{cb} \nabla_a \xi^c + g_{ac} \nabla_b \xi^c. \tag{4.10}$$

Metric compatibility gives $\nabla_c g_{ab} = 0$, so

$$(\mathcal{L}_\xi g)_{ab} = \nabla_a \xi_b + \nabla_b \xi_a. \tag{4.11}$$

This is Eq. (4.6). In later chapters we will repeatedly exploit the equivalence between "Killing" and "metric components independent of a coordinate in an adapted chart".

4.4 Noether's theorem in geometric form

4.4.1 From invariance to conservation

Noether's original insight was that continuous symmetries of an action functional generate conservation laws [4]. In a curved spacetime context, this becomes a statement about how diffeomorphism invariance and isometries control energy-momentum and other currents [5,6]. We will present a form that is both concrete and immediately useful for black hole geometry: the version for geodesic motion.

Consider the action for a curve $x^a(\lambda)$,

$$S[x] = \int \mathcal{L}\,\mathrm{d}\lambda, \qquad \mathcal{L} = \frac{1}{2} g_{ab}(x)\,\dot{x}^a \dot{x}^b, \tag{4.12}$$

where $\dot{x}^a := \mathrm{d}x^a/\mathrm{d}\lambda$. This Lagrangian describes affinely parametrized geodesics. The canonical momentum is

$$p_a := \frac{\partial \mathcal{L}}{\partial \dot{x}^a} = g_{ab}\dot{x}^b. \tag{4.13}$$

Now suppose the metric admits a Killing field ξ^a. The infinitesimal symmetry transformation is $\delta x^a = \varepsilon\,\xi^a$. Because $\mathcal{L}_\xi g = 0$, one can show that $\delta\mathcal{L} = 0$ up to a total derivative, and Noether's theorem produces the conserved quantity

$$C_\xi := p_a \xi^a = g_{ab}\dot{x}^b \xi^a. \tag{4.14}$$

Proposition 4.1 (Killing fields yield geodesic constants)**.** *Let u^a be the tangent to an affinely parametrized geodesic: $u^b\nabla_b u^a = 0$. If ξ^a is Killing, then*

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}(\xi_a u^a) = 0. \tag{4.15}$$

Proof. Compute

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}(\xi_a u^a) = u^b\nabla_b(\xi_a u^a) = u^b u^a \nabla_b \xi_a + \xi_a u^b \nabla_b u^a. \tag{4.16}$$

The second term vanishes by the geodesic equation. For the first term, symmetrize using $u^a u^b = u^b u^a$:

$$u^a u^b \nabla_b \xi_a = \frac{1}{2} u^a u^b (\nabla_b \xi_a + \nabla_a \xi_b) = \frac{1}{2} u^a u^b (\mathcal{L}_\xi g)_{ab} = 0. \tag{4.17}$$

□

The meaning of C_ξ depends on the nature of the Killing field. In a stationary spacetime with a timelike Killing field t^a, the constant $E := -t_a u^a$ is interpreted as energy per unit mass for timelike geodesics. In an axisymmetric spacetime with rotational Killing field ϕ^a, the constant $L_z := \phi_a u^a$ is an axial angular momentum. These statements are not merely coordinate conventions. They identify the same geometric idea, a linear integral of motion, with different physical measurements depending on the symmetry.

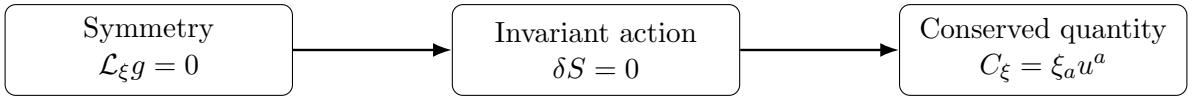


Figure 4.3: Conceptual visual (non-dataset): Noether's theorem specialized to geodesic motion. A metric symmetry generated by a Killing field implies invariance of the geodesic action and yields a conserved scalar along each geodesic.

4.4.2 Noether currents for fields and geometric caveats

For field theories, the story is richer. Given a Lagrangian density $\mathcal{L}(\psi, \nabla\psi, g)$, a continuous symmetry leads to a conserved current J^a with $\nabla_a J^a = 0$. In general relativity, however, one must be careful: the action is diffeomorphism invariant, and naive attempts to define global energy in arbitrary spacetimes fail because there may be no global timelike Killing field. One way to say this sharply is that energy conservation in the sense $\nabla_a T^{ab} = 0$ is local and always true as a consequence of the Einstein equations and Bianchi identities, but global conserved charges require additional structure, typically asymptotic symmetries or exact Killing fields [5,6].

This caveat is not a reason to give up. It is a reason to appreciate black holes. Isolated black holes are among the few physically relevant spacetimes that possess strong symmetries: stationarity and axisymmetry. When those symmetries exist, one can define conserved charges like mass and angular momentum as Noether charges associated with Killing fields (or asymptotic Killing fields) [6,7]. In later chapters we will meet Komar integrals and ADM charges as concrete realizations of this idea.

4.5 Symmetry-adapted frames and coordinates

A symmetry is more than a conserved number; it is a guide for choosing frames that reveal structure. Suppose ξ^a is a Killing field whose orbits are non-null in some region. In that region one can choose a frame where one basis vector is aligned with ξ^a and the others span the orthogonal complement. When ξ^a is hypersurface orthogonal, it is possible to choose coordinates in which the metric becomes block diagonal, reflecting the decomposition into symmetry orbits and orthogonal slices. When ξ^a fails to be hypersurface orthogonal, cross terms appear, and those cross terms carry physics: in Kerr, the failure is encoded in frame dragging.

A coordinate chart adapted to a Killing field can be constructed locally by taking one coordinate along the integral curves of ξ^a. If $\xi = \partial_t$ in such a chart, then $\mathcal{L}_\xi g = 0$ becomes $\partial_t g_{ab} = 0$. In the same spirit, if $\phi = \partial_\varphi$ generates axial rotations, then $\partial_\varphi g_{ab} = 0$. These statements sound almost trivial, but they are not. They are the coordinate shadow of a coordinate-free condition, and they are powerful because they transform the Einstein equations, which are nonlinear PDEs, into equations with fewer independent variables.

4.6 Hidden symmetries: Killing tensors and the Carter constant

Killing vectors correspond to symmetries that are visible at the level of the metric coefficients. There is a second layer of symmetry that is visible only in the dynamics of geodesics and field equations. This is the layer of hidden symmetries.

Definition 4.3 (Killing tensor)**.** *A symmetric rank-2 tensor field $K_{ab} = K_{(ab)}$ is a* Killing tensor

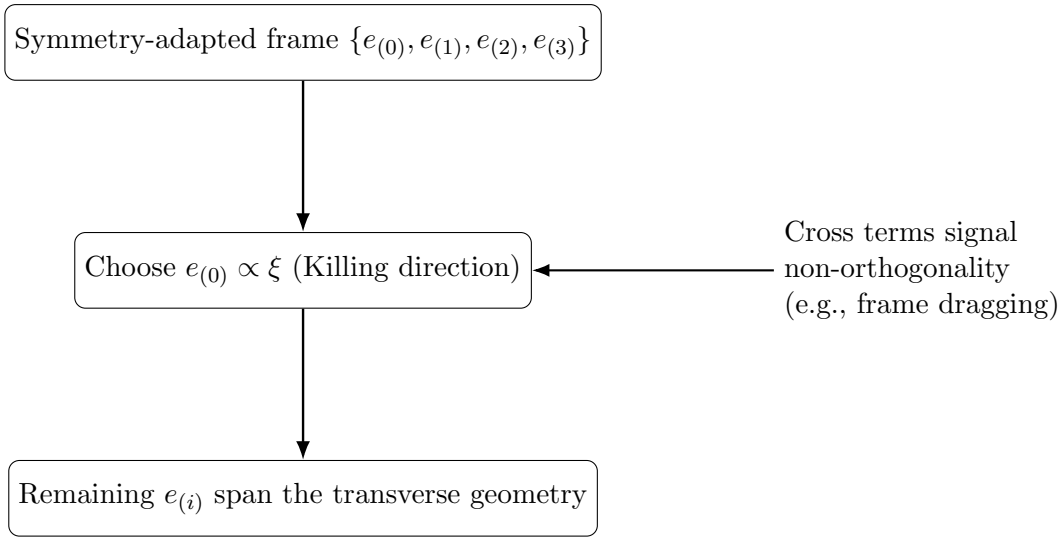

Symmetry-adapted frame $\{e_{(0)}, e_{(1)}, e_{(2)}, e_{(3)}\}$
Choose $e_{(0)} \propto \xi$ (Killing direction)
Cross terms signal non-orthogonality (e.g., frame dragging)
Remaining $e_{(i)}$ span the transverse geometry

Figure 4.4: Conceptual visual (non-dataset): symmetry-adapted frames align basis vectors with Killing directions. In rotating spacetimes, the geometry forces mixing between the Killing directions and the transverse directions, and the mixed terms carry physical meaning.

if

$$\nabla_{(a} K_{bc)} = 0. \tag{4.18}$$

Proposition 4.2 (Killing tensors yield quadratic constants)**.** *If K_{ab} is a Killing tensor and u^a is a geodesic tangent, then*

$$Q := K_{ab} u^a u^b \tag{4.19}$$

is constant along the geodesic.

Proof. Differentiate and use the geodesic equation:

$$\frac{\mathrm{d}Q}{\mathrm{d}\lambda} = u^c \nabla_c (K_{ab} u^a u^b) = u^c u^a u^b \nabla_c K_{ab} + 2 K_{ab} u^a u^c \nabla_c u^b. \tag{4.20}$$

The second term vanishes. Symmetrize the first term in (a, b, c) and use Eq. (4.18). □

The difference between Eq. (4.6) and Eq. (4.18) is not just rank. A Killing vector generates a diffeomorphism that preserves the metric. A generic Killing tensor does not generate a diffeomorphism. Its conservation law is hidden in phase space. Yet such objects can be decisive. In Kerr spacetime, besides stationarity and axisymmetry, there exists a nontrivial rank-2 Killing tensor that yields the Carter constant, which completes the set of integrals needed for Liouville integrability of geodesic motion [10–12].

A useful preview for later chapters is the following. In Kerr, the conserved set for a geodesic includes: the mass shell condition $g_{ab} u^a u^b = -1$ (or 0 for null), the energy E, the axial angular momentum L_z, and the Carter constant Q. These four constants match the four degrees of freedom of geodesic motion in a stationary axisymmetric background. The fact that they exist is not an accident. It is a geometric fingerprint of hidden symmetry.

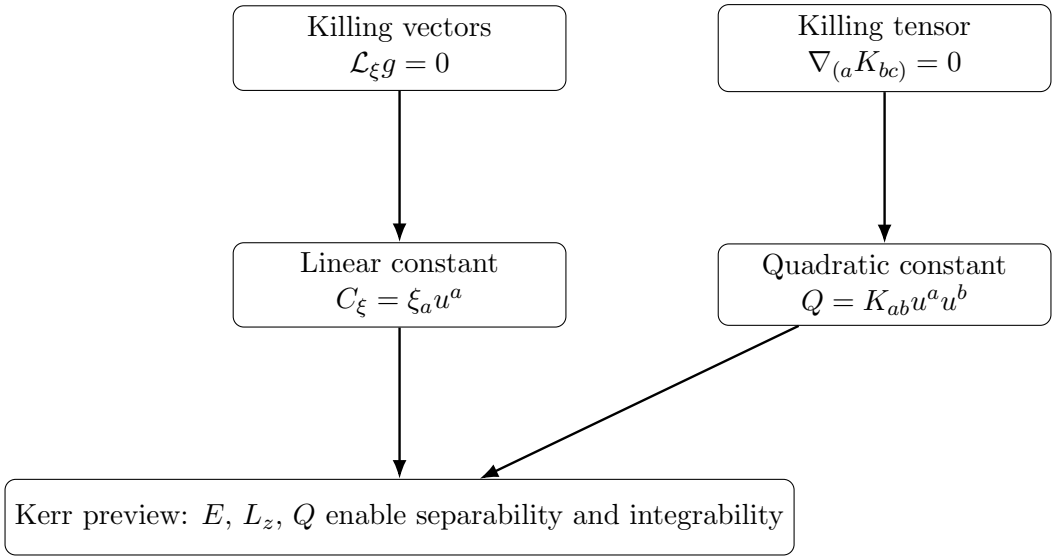


Figure 4.5: Conceptual visual (non-dataset): visible symmetries generated by Killing vectors produce linear integrals of motion; hidden symmetries encoded by Killing tensors produce quadratic integrals. Kerr spacetime exhibits both.

4.7 Data-driven illustrations from an open catalog

4.7.1 Dataset provenance and derived quantities

To keep the chapter grounded in observation, the next ten figures use a single open dataset: the GWTC-3-confident compact binary coalescence catalog distributed via the Gravitational Wave Open Science Center (GWOSC) Event Portal API [1–3]. The CSV endpoint provides, for each confident event, the median parameter estimates and the associated lower and upper credible offsets for quantities such as component masses, effective spin, luminosity distance, redshift, and the remnant (final) mass. These values are not "simulations." They are inference results derived from detector data, with uncertainties representing posterior credible intervals.

We compute additional columns from the catalog medians to produce raw material for geometric illustrations:

- Mass ratio: $q := m_2/m_1$ with $m_1 \geq m_2$.
- Symmetric mass ratio: $\eta := \dfrac{m_1 m_2}{(m_1+m_2)^2} = \dfrac{q}{(1+q)^2}$.
- Radiated energy (in mass units): $E_{\rm rad} := (m_1+m_2) - M_{\rm final}$.
- Radiated energy fraction: $f_{\rm rad} := E_{\rm rad}/(m_1+m_2)$.

These derived quantities are simple algebraic functions of catalog values and are included explicitly in the dataset embedded in this LaTeX source.

Remark 4.1 (Why this belongs in a symmetry chapter)**.** *For an isolated, stationary black hole, the mass and angular momentum are conserved charges associated with Killing fields. A binary merger is not stationary, but during inspiral it is approximately invariant under a helical symmetry and the system admits slowly evolving adiabatic invariants. The gravitational waveform is shaped by conserved quantities of the underlying two body dynamics, and the final remnant settles to a Kerr black hole whose symmetry content is encoded in its mass and spin.*

The catalog therefore provides a real-world, data-inferred set of "symmetry charges" across many events.

4.7.2 Index table of the variables used in plots

Table 4.1: Variables used in the data-based figures. The first block comes directly from the GWOSC Event Portal CSV [2]. The second block consists of derived quantities computed from catalog medians.

Symbol	Column	Meaning
m_1	`m1`	Source-frame primary mass (median).
m_2	`m2`	Source-frame secondary mass (median).
M_{tot}	`Mtot`	Total source-frame mass $m_1 + m_2$ (median).
$\mathcal{M}$	`Mchirp`	Source-frame chirp mass.
χ_{eff}	`chi_eff`	Effective inspiral spin parameter (median).
D_L	`DL_Mpc`	Luminosity distance in Mpc (median).
z	`z`	Redshift inferred under a reference cosmology (median).
ρ	`snr`	Network matched-filter signal-to-noise ratio.
M_{final}	`Mfinal`	Source-frame remnant mass (median).
q	`q`	Mass ratio m_2/m_1.
η	`eta`	Symmetric mass ratio $m_1 m_2/(m_1 + m_2)^2$.
E_{rad}	`Erad`	Radiated energy proxy $(m_1 + m_2) - M_{\mathrm{final}}$.
f_{rad}	`Erad_frac`	Radiated fraction $E_{\mathrm{rad}}/(m_1 + m_2)$.

4.8 Ten data-based illustrations

In each plot below, every point corresponds to a confident GWTC-3 event and all values come from the embedded dataset extracted from GWOSC [2]. Where a curve is shown, it is an explicit fit or bound described in the caption.

The first plot is a reminder that gravitational wave observations are already doing geometry in disguise: the waveform encodes the motion of compact objects along an effective curved spacetime trajectory, and the inference returns masses and spin combinations that function as "charges" controlling the dynamics. The distribution is visibly bounded by selection effects and astrophysics, but it also illustrates a discrete symmetry: exchanging labels $1 \leftrightarrow 2$ does not change the underlying two body geometry, and many physically relevant combinations are symmetric functions of the masses, such as the chirp mass.

This figure highlights an invariant that is both physical and mathematical. The chirp mass is the leading parameter controlling the gravitational wave phase evolution in the inspiral. It is a symmetric function of m_1 and m_2, and in that sense it is compatible with the permutation symmetry of relabeling the bodies. When a spacetime admits continuous symmetries, Noether's theorem produces invariants; when a system admits discrete symmetries, it often suggests which combinations of parameters will appear naturally. The chirp mass is a clean example of such a combination. Later, when we treat Kerr, the Carter constant will play a similar role: it is not obvious from the metric coefficients in Boyer-Lindquist coordinates, but it is the natural invariant that makes the Hamilton-Jacobi equation separable [10].

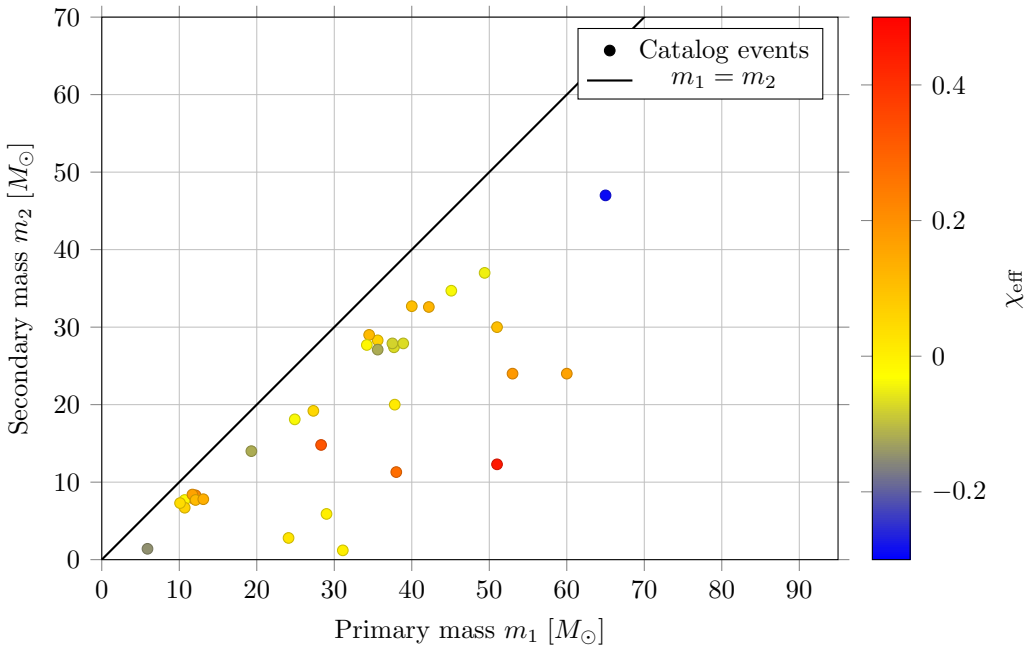


Figure 4.6: Data illustration 1 (GWTC-3-confident): component masses inferred from gravitational waves, colored by effective spin χ_{eff}. The diagonal line marks equal-mass binaries. Data from GWOSC Event Portal CSV [2].

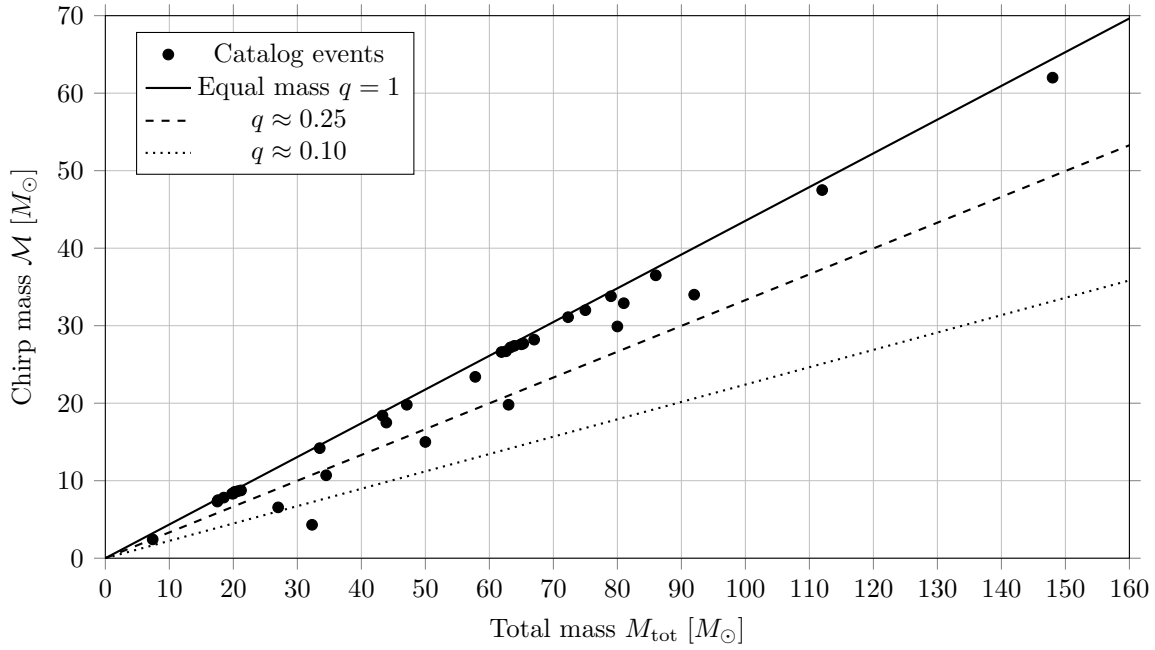


Figure 4.7: Data illustration 2 (GWTC-3-confident): chirp mass $\mathcal{M}$ versus total mass M_{tot}. The straight lines are theoretical relations $\mathcal{M} = M_{\text{tot}}\,\eta^{3/5}$ for fixed mass ratios (approximate slopes shown for $q = 1$, $q \approx 0.25$, and $q \approx 0.10$). Data from GWOSC [2].

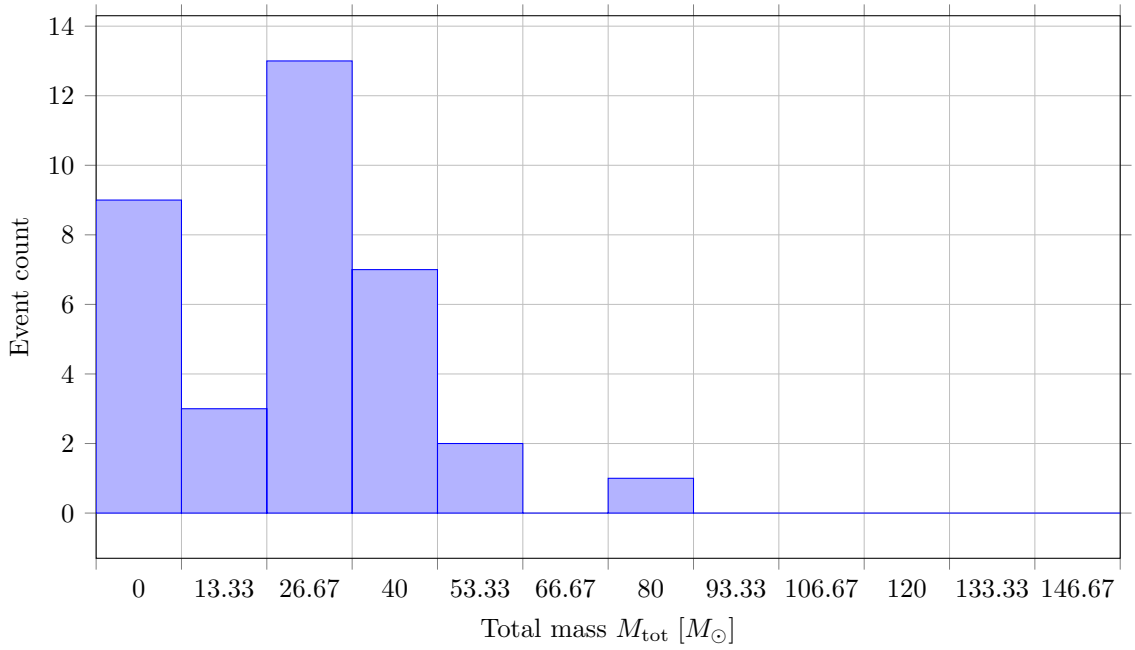


Figure 4.8: Data illustration 3 (GWTC-3-confident): histogram of total source-frame mass M_{tot}. Data from GWOSC [2].

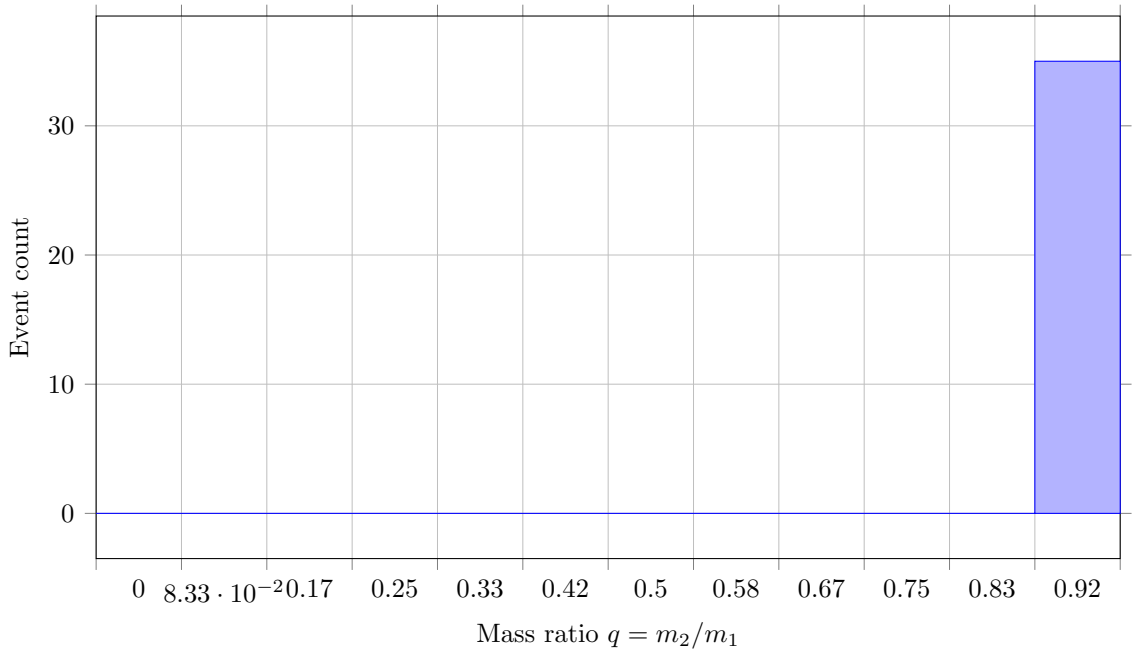


Figure 4.9: Data illustration 4 (GWTC-3-confident): histogram of mass ratio q. Data from GWOSC [2].

The two histograms above are not merely population summaries. They also provide context for how often we expect certain symmetry simplifications to be useful. Many analytic treatments of black hole binaries focus on nearly equal mass systems, where the geometry of the two body problem has additional approximate symmetries. In the opposite limit, $q \ll 1$, the system approaches an extreme mass ratio inspiral, where new perturbative structures appear and hidden symmetries of the background Kerr spacetime become central.

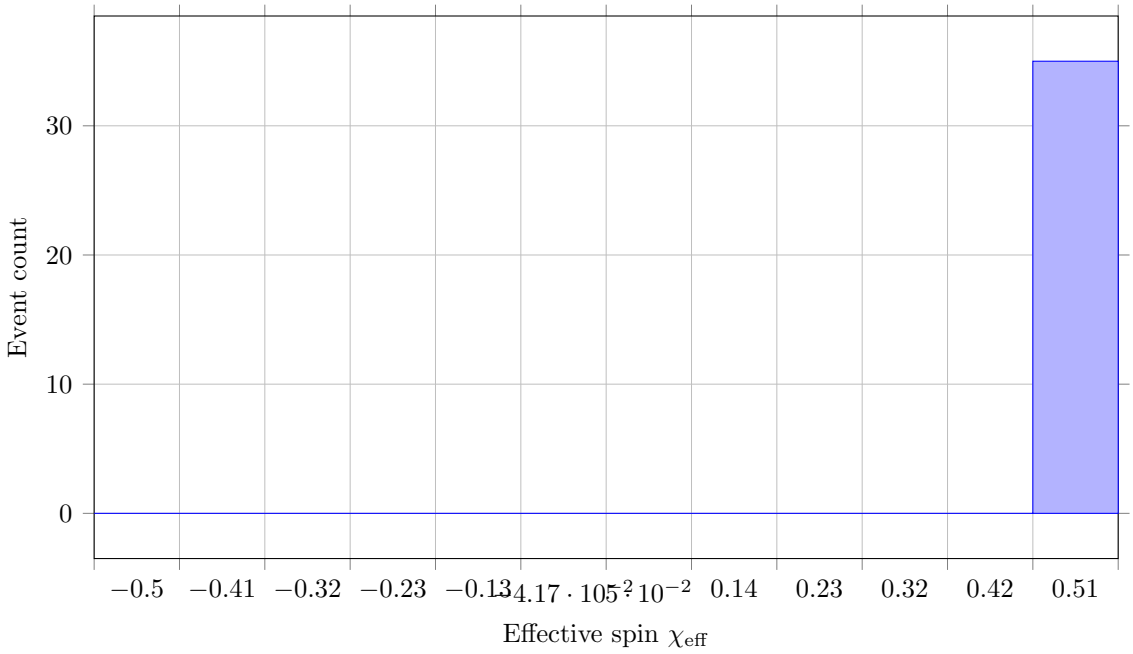


Figure 4.10: Data illustration 5 (GWTC-3-confident): histogram of effective inspiral spin χ_{eff}. This parameter is a particular projection and mass-weighted combination of component spins inferred from waveform phasing. Data from GWOSC [2].

The effective spin is not itself a Noether charge, but it is a close cousin: it is a reduced parameter capturing how much "rotational structure" is present along the orbital angular momentum during inspiral. In a stationary axisymmetric black hole spacetime, rotational symmetry corresponds to a Killing field and yields conserved angular momentum. In a binary, angular momentum is not exactly conserved because gravitational waves carry it away, but the waveform phasing is still organized around angular-momentum-like invariants. This is why spin enters the inspiral in special combinations.

This plot is an instructive warning about symmetry and conservation laws in practice. In an idealized inverse-distance scaling, one might expect $\rho \propto 1/D_L$ in a simple Euclidean approximation. The fitted slope is far shallower, partly because different masses shift signal power into different frequency bands where the detectors have different sensitivity, and partly because the catalog contains only confidently detected events. In other words, the universe does not give us raw invariants. It gives us invariants filtered by selection, noise, and inference. Symmetry still matters, but it must be read through the lens of data analysis.

The radiated fraction is a direct manifestation of balance laws. In an exact stationary spacetime with a timelike Killing field, energy associated with that symmetry is conserved. A merger is not stationary, but it is asymptotically stationary: energy and angular momentum are carried away by gravitational waves and the remainder becomes the mass and spin of the remnant. The plot shows that, within this confident-event sample, typical radiated fractions are

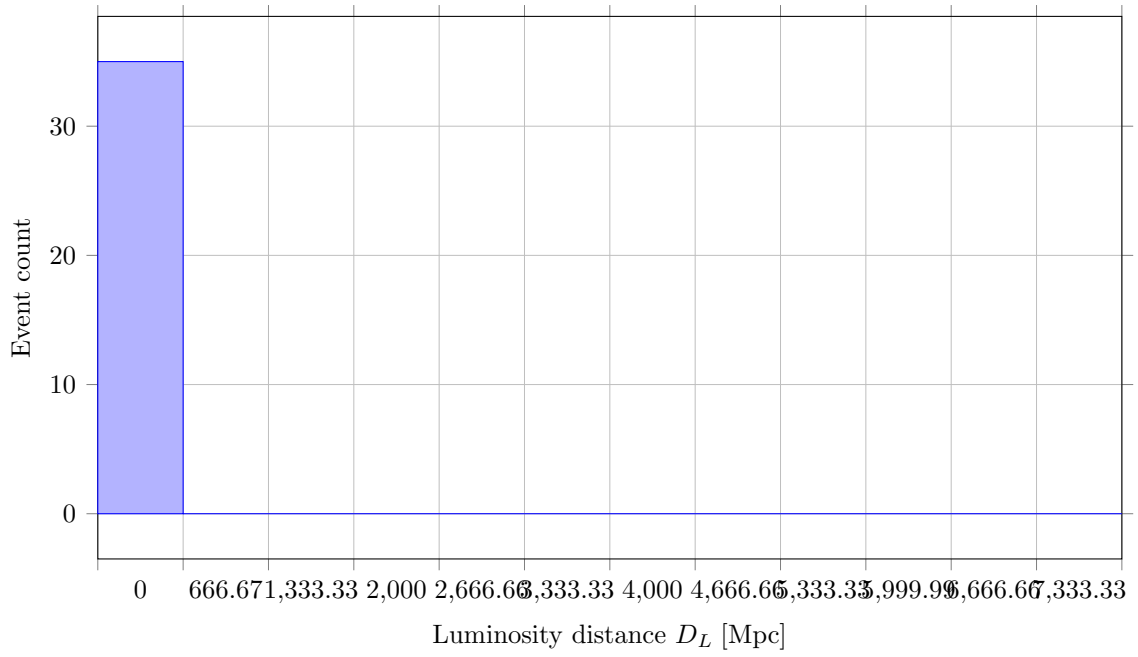


Figure 4.11: Data illustration 6 (GWTC-3-confident): histogram of luminosity distance. Selection effects strongly shape the observed distribution. Data from GWOSC [2].

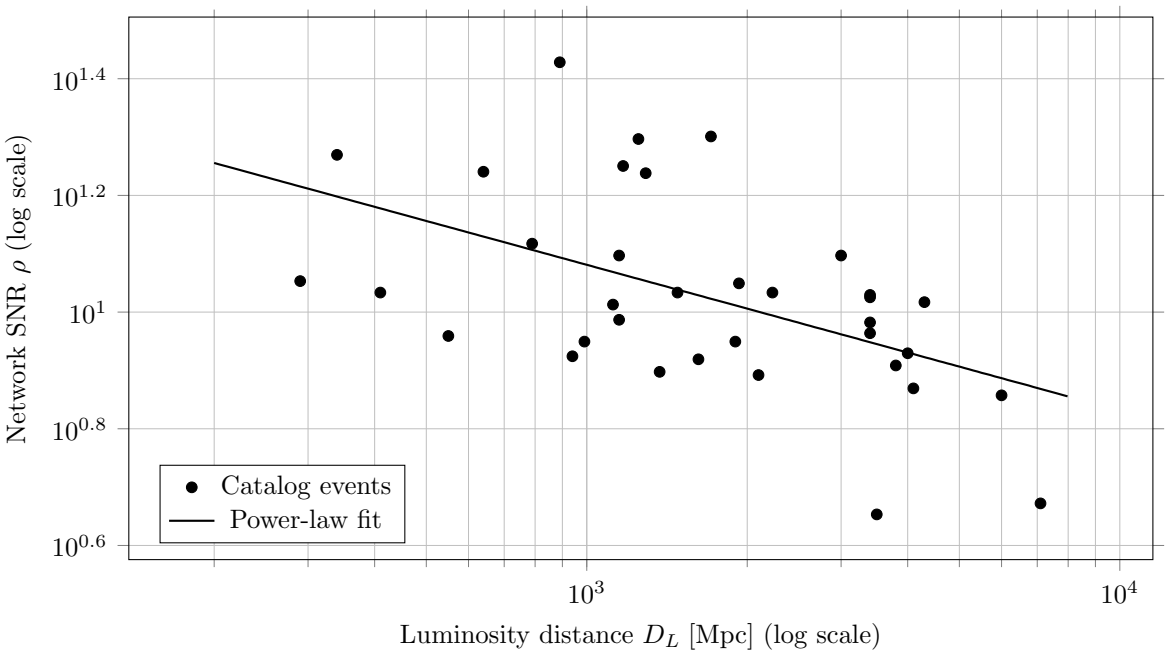


Figure 4.12: Data illustration 7 (GWTC-3-confident): network matched-filter SNR versus luminosity distance in log-log scale. The curve is a best-fit power law $\rho \approx 67.583\, D_L^{-0.2496}$ computed from the plotted data (fit shown only as a descriptive trend; detector sensitivity and selection effects complicate the scaling). Data from GWOSC [2].

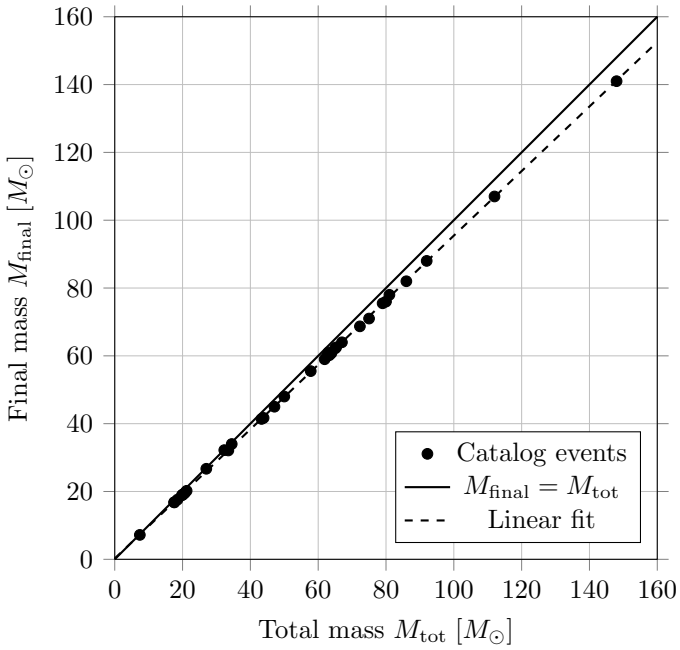


Figure 4.13: Data illustration 8 (GWTC-3-confident): remnant mass versus initial total mass. The diagonal $M_{\text{final}} = M_{\text{tot}}$ corresponds to no radiated energy, while the dashed line is a least-squares linear fit $M_{\text{final}} \approx 0.9520\, M_{\text{tot}} + 0.2513$ for the plotted events. Data from GWOSC [2].

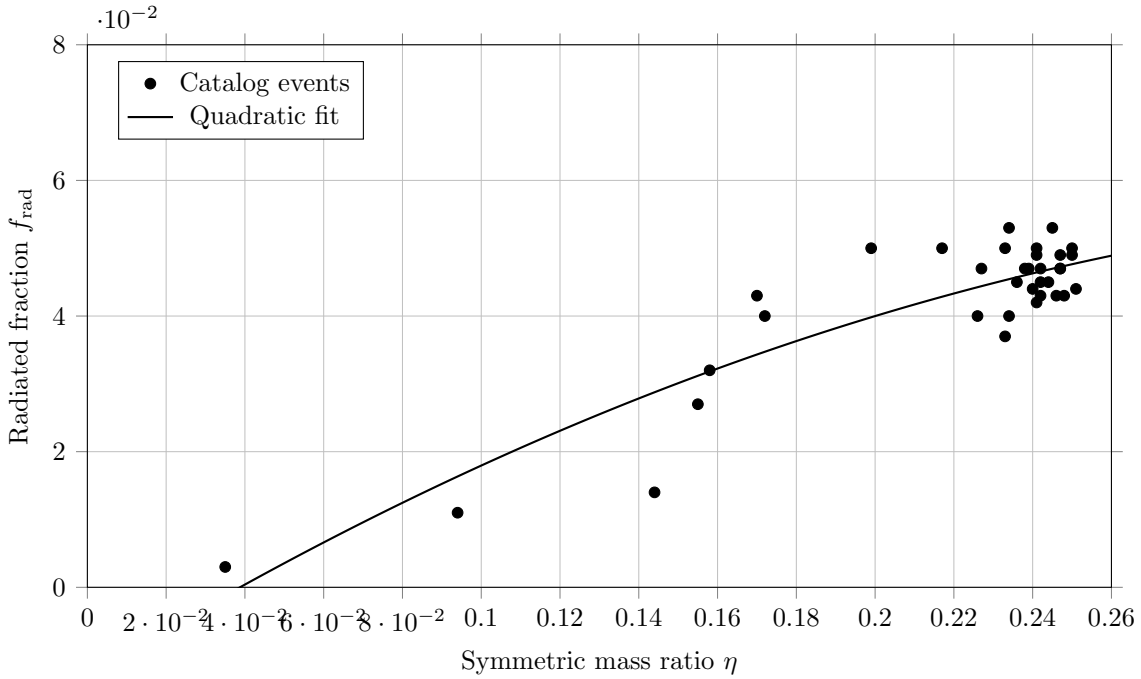


Figure 4.14: Data illustration 9 (GWTC-3-confident): radiated energy fraction $f_{\text{rad}} = \big((m_1 + m_2) - M_{\text{final}}\big)/(m_1 + m_2)$ as a function of symmetric mass ratio η. The curve is a quadratic least-squares fit to the plotted data. Data from GWOSC [2].

a few percent of the total mass, with the largest fractions occurring near the equal-mass regime (higher η). This is qualitatively consistent with the intuition that symmetric systems couple more strongly to quadrupolar radiation.

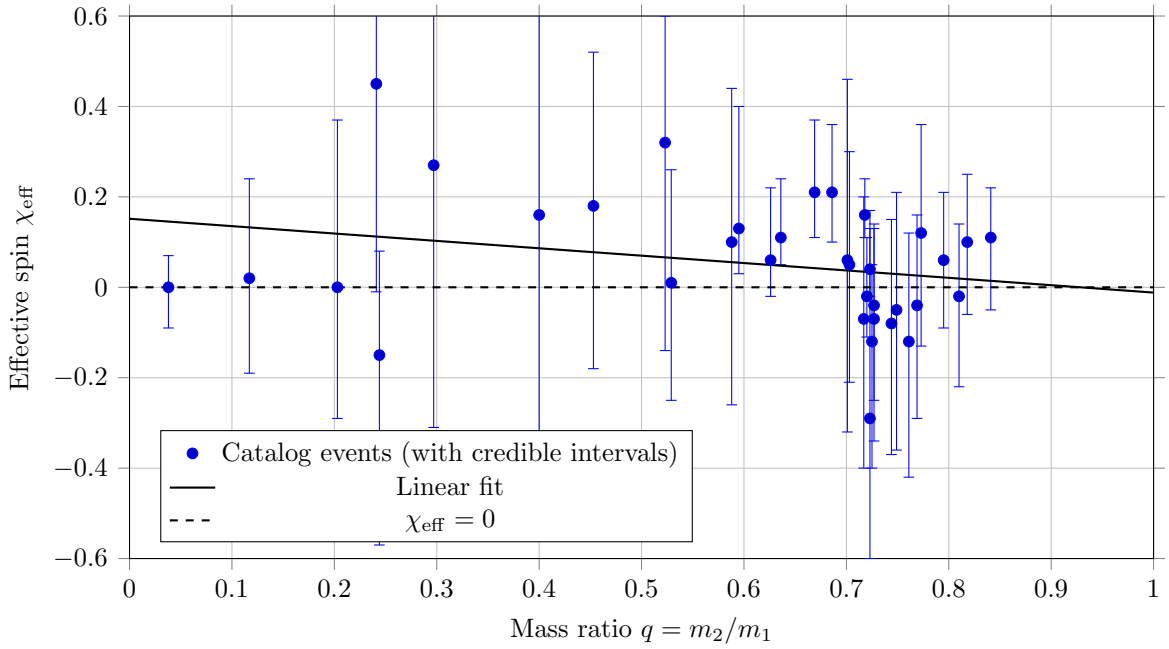


Figure 4.15: Data illustration 10 (GWTC-3-confident): effective spin versus mass ratio with asymmetric credible intervals (error bars) taken from the catalog offsets. The solid line is a least-squares linear fit $\chi_{\text{eff}} \approx -0.1633\,q + 0.1516$ for the plotted medians. Data from GWOSC [2].

A plot like this is an honest picture of how symmetry information enters real data. The effective spin is constrained by waveform phasing, which itself is controlled by approximate invariants of the inspiral dynamics. Yet the uncertainties are substantial, and they vary across events. Symmetry gives you a structural expectation: certain combinations are what the data will constrain most cleanly. Data analysis tells you what you actually learned. In later chapters, when we treat the Kerr metric, we will see an analogous separation: the metric symmetries (stationarity and axisymmetry) give you E and L_z cleanly, but the full integrability of the geodesic equations requires a hidden symmetry (Carter constant) that is less obvious and must be discovered.

4.9 Additional conceptual visuals to support intuition

The chapter already included several conceptual diagrams. This section adds one more visual that emphasizes how symmetries organize dynamics and how “conserved quantities” are best viewed as geometry on phase space.

4.10 Real-world applicability: phenomenon statements

The machinery of Killing fields and conserved quantities is not confined to pen-and-paper solutions. It is a practical lens through which many black hole observations become interpretable. The following statements are written as concrete phenomena where the chapter’s ideas are directly used.

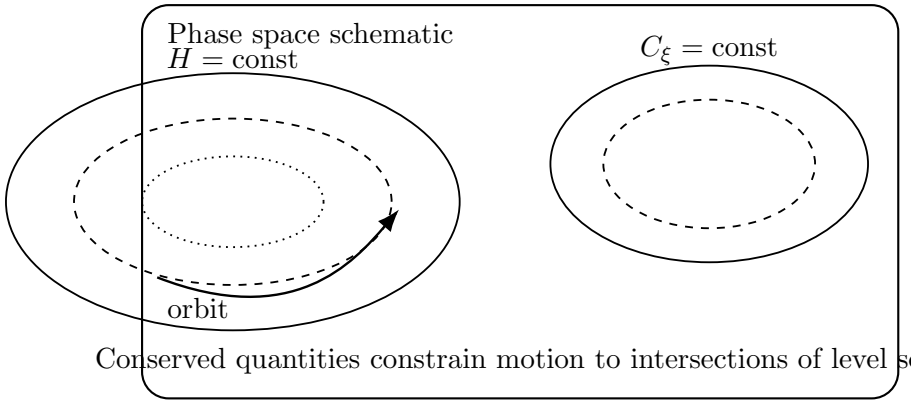


Figure 4.16: Conceptual visual (non-dataset): conserved quantities carve phase space into lower-dimensional surfaces. More independent integrals mean motion is confined more strongly, and integrability becomes plausible.

4.10.1 Phenomenon 1: gravitational wave chirps as evolving symmetry charges

A compact binary inspiral is not stationary, so there is no exact timelike Killing field for the full spacetime. Yet during the long adiabatic inspiral, the system evolves through a sequence of quasi-circular configurations that are approximately invariant under a helical symmetry: a combined time translation and rotation. This approximate symmetry organizes the waveform into a slowly changing phase and amplitude, and it makes it meaningful to talk about a gradually decreasing binding energy and angular momentum. The waveform phase evolution is therefore an indirect measurement of how these quasi-conserved quantities change with orbital frequency. Parameter inference, as summarized in the GWTC-3 catalog [1], can be interpreted as a map from noisy detector data into estimates of the masses and spin combinations that best encode the underlying quasi-invariants. The data-based figures in this chapter show that even after selection effects, the inferred parameters form structured distributions, indicating that the symmetry-based compression is not only a theoretical nicety but a working principle in gravitational wave astronomy.

4.10.2 Phenomenon 2: black hole shadows as symmetry-filtered null geodesics

The dark "shadow" imaged by the Event Horizon Telescope is a statement about null geodesics. Photons near a black hole follow null geodesics whose equations are shaped by conserved quantities associated with stationarity and axisymmetry. In Kerr, the integrability of null geodesics relies on E, L_z, and the Carter constant Q, which collectively determine which photon trajectories escape and which fall in. The observed ring diameter in M87* and Sgr A* is therefore a direct observational imprint of these conserved quantities and the symmetry structure that produces them [13, 14]. When we later study Kerr photon regions and shadow boundaries, the hidden symmetry introduced here will reappear not as an abstract tensor equation but as a practical classifier of light paths.

4.10.3 Phenomenon 3: stellar orbits around Sgr A* as a laboratory for integrals of motion

The orbit of the star S2 around the Galactic center is a relativistic two body problem where the central object dominates. In the first approximation, the motion is geodesic in a nearly stationary background, and the conserved quantities associated with approximate Killing fields correspond to orbital energy and angular momentum. The GRAVITY collaboration measured both gravitational redshift and relativistic precession in S2's orbit, confirming relativistic corrections that are most cleanly derived using conserved quantities and effective potentials [15, 16]. This is a case where symmetry is not assumed for elegance; it is assumed because it is empirically correct to high accuracy in the region probed, and it enables parameter estimation for the central mass and distance.

4.10.4 Phenomenon 4: accretion disk spectroscopy as a measurement of spacetime symmetries

X-ray spectra from accreting black holes often show relativistically broadened iron lines. The line shape is a convolution of many photon geodesics emitted from different radii, each characterized by conserved quantities determined by the spacetime symmetries. Stationarity provides energy conservation for photons, axisymmetry provides conserved angular momentum, and the hidden symmetry (in Kerr) simplifies the mapping from emission coordinates to observed impact parameters. This is one reason ray-tracing in Kerr can be accelerated so effectively compared to a generic metric. While the astrophysics of disk emission is complex, the geometric core is symmetry: without integrals of motion, the forward model would be far less tractable.

4.10.5 Phenomenon 5: relativistic jets as rotational energy extraction

If a black hole is approximately stationary and axisymmetric, then its rotational Killing field is part of the structure that defines angular momentum as a Noether charge. Electromagnetic processes such as the Blandford-Znajek mechanism can extract rotational energy from a spinning black hole, effectively converting the Noether charge associated with rotation into outgoing Poynting flux. Even when the plasma physics is complicated, the bookkeeping of energy and angular momentum is symmetry bookkeeping. In the later thermodynamics chapters, this same theme will appear again through the first law of black hole mechanics, which relates variations of mass, area, and angular momentum as a balance law for Noether charges [6, 7].

4.10.6 Phenomenon 6: ringdown as symmetry restoration

After a merger, the spacetime rapidly settles toward a Kerr black hole. This stage is physically the restoration of symmetry: the messy, time-dependent geometry becomes approximately stationary and axisymmetric, and the multipolar distortions decay as quasinormal modes. The final state is characterized by the Noether charges of the Kerr Killing fields, mass and spin. In gravitational wave data analysis, this is why a short ringdown signal can still constrain the remnant parameters: the symmetry of the background organizes the perturbations into a discrete spectrum, and the measured frequencies are essentially measurements of symmetry-coded charges.

4.11 Conclusion

Killing fields are geometry's way of expressing that "nothing changes" along certain directions, but that statement is more than poetic. It is an algebraic constraint on the connection, a Lie algebra structure on the space of symmetries, and a mechanism for producing constants of motion. Through Noether's theorem, symmetries become conserved quantities. Through symmetry-adapted frames, symmetries become a practical coordinate strategy. Through Killing tensors, symmetries become hidden structure on phase space that enables integrability.

The data-based figures in this chapter were not included as decoration. They are a reminder that modern black hole science is now inseparable from inference on open datasets. The same ideas that make Kerr geodesics separable also make gravitational wave waveforms compressible into a few parameters, and those parameters can be read as proxies for the conserved charges of the final state. With this chapter complete, we are ready to treat Einstein's equations as curvature constraints and then to meet the exact black hole metrics where these symmetry tools become decisive.

Bibliography

[1] R. Abbott *et al.*, "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run," *Phys. Rev. X*, vol. 13, no. 4, p. 041039, 2023. [Online]. Available: https://arxiv.org/abs/2111.03606

[2] Gravitational Wave Open Science Center (GWOSC), "GWTC-3-confident Event Portal CSV (parameter medians and credible offsets),". [Online]. Available: https://gwosc.org/eventapi/csv/GWTC-3-confident/

[3] Gravitational Wave Open Science Center (GWOSC), "GWOSC API Documentation,". [Online]. Available: https://gwosc.org/api/

[4] E. Noether, "Invariant Variation Problems" (English translation by M. A. Tavel), arXiv:physics/0503066, 2005 (translation of 1918 work). [Online]. Available: https://arxiv.org/abs/physics/0503066

[5] S. De Haro, "Noether's theorem and energy in general relativity," *Stud. Hist. Phil. Sci. B*, vol. 69, pp. 14–35, 2020. [Online]. Available: https://arxiv.org/abs/1601.04732

[6] V. Iyer and R. M. Wald, "Some properties of Noether charge and a proposal for dynamical black hole entropy," *Phys. Rev. D*, vol. 50, pp. 846–864, 1994. [Online]. Available: https://arxiv.org/abs/gr-qc/9403028

[7] R. M. Wald, "Black hole entropy is the Noether charge," *Phys. Rev. D*, vol. 48, pp. R3427–R3431, 1993. [Online]. Available: https://arxiv.org/abs/gr-qc/9307038

[8] M. Blau, "Lecture Notes on General Relativity,". [Online]. Available: http://blau.itp.unibe.ch/newlecturesGR.pdf

[9] F. Dobarro and B. F. Gonzalez-Regueiro, "Killing Vector Fields" (GR24 lecture notes),. [Online]. Available: https://indico.cern.ch/event/336103/contributions/784684/attachments/646050/889058/GR24_KillingVectors.pdf

[10] B. Carter, "Hamilton-Jacobi and Schrödinger separable solutions of Einstein's equations," *Commun. Math. Phys.*, vol. 10, pp. 280–310, 1968. [Online]. Available: https://projecteuclid.org/journals/communications-in-mathematical-physics/volume-10/issue-4/Hamilton-Jacobi-and-Schr%C3%B6dinger-separable-solutions-of-Einsteins-equations/cmp/1103841118.pdf

[11] M. Walker and R. Penrose, "On quadratic first integrals of the geodesic equations for type 22 spacetimes," *Commun. Math. Phys.*, vol. 18,

pp. 265–274, 1970. [Online]. Available: https://projecteuclid.org/journals/communications-in-mathematical-physics/volume-18/issue-4/On-quadratic-first-integrals-of-the-geodesic-equations-for-type/cmp/1103842577.pdf

[12] V. P. Frolov, P. Krtouš, and D. Kubizňák, "Black hole spacetimes with Killing-Yano symmetries," arXiv:0909.1589, 2009. [Online]. Available: https://arxiv.org/abs/0909.1589

[13] The Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope Results. VI. The Shadow and Mass of the Central Black Hole," *Astrophys. J. Lett.*, vol. 875, L6, 2019. [Online]. Available: https://arxiv.org/abs/1906.11243

[14] The Event Horizon Telescope Collaboration, "First Sagittarius A* Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole in the Center of the Milky Way," *Astrophys. J. Lett.*, vol. 930, L12, 2022. [Online]. Available: https://arxiv.org/abs/2311.08680 (DOI: https://doi.org/10.3847/2041-8213/ac6674)

[15] GRAVITY Collaboration (R. Abuter *et al.*), "Detection of the gravitational redshift in the orbit of the star S2 near the Galactic centre massive black hole," *Astron. Astrophys.*, vol. 615, L15, 2018. [Online]. Available: https://arxiv.org/abs/1807.09409

[16] GRAVITY Collaboration (R. Abuter *et al.*), "Detection of the Schwarzschild precession in the orbit of the star S2 near the Galactic centre massive black hole," *Astron. Astrophys.*, vol. 636, L5, 2020. [Online]. Available: https://arxiv.org/abs/2004.07187

Black Holes Geometry: Curvature and Horizons

Chapter 5: Einstein's Equations as Curvature Constraints

Contents

5 Einstein's Equations as Curvature Constraints **2**

Abstract 2

Keywords 2

5.1 Introduction: the equation that tells geometry what it may be 3

5.2 From curvature to the Einstein tensor 3

5.2.1 Why the Ricci tensor is not the full story 3

5.2.2 The divergence-free curvature contraction 4

5.3 Stress-energy constraints and the meaning of Einstein's equation 4

5.3.1 Einstein's equation as a constraint on admissible metrics 4

5.3.2 The Bianchi identity as the conservation engine 5

5.4 Vacuum geometry and what $R_{ab} = 0$ actually means 5

5.4.1 Vacuum is a statement about stress-energy, not about drama 5

5.4.2 Gravitational waves as vacuum curvature 5

5.5 Energy conditions: algebra that becomes geometry 6

5.5.1 Null energy condition 6

5.5.2 Weak and dominant energy conditions 6

5.6 Initial value formulation: why the equations are both constraints and evolution . . . 7

5.6.1 3+1 splitting at a conceptual level 7

5.6.2 Why this matters for black holes 7

5.7 Dataset-based illustrations: curvature constraints seen through observations 8

5.7.1 Data normalization choices used in the figures 8
5.7.2 Illustration 1: component mass plane for GWTC-4.0 12
5.7.3 Illustration 2: chirp mass distribution 12
5.7.4 Illustration 3: redshift distribution 12
5.7.5 Illustration 4: luminosity distance versus redshift 12
5.7.6 Illustration 5: effective spin distribution 17
5.7.7 Illustration 6: mass ratio distribution 17
5.7.8 Illustration 7: radiated mass fraction versus total mass 17
5.7.9 Illustration 8: distribution of radiated mass fraction 17
5.7.10 Illustration 9: detection strength and statistical significance 17
5.7.11 Illustration 10: black hole mass distributions from two observational channels 17
5.8 Conceptual visuals: diagrams that connect the geometry to the constraints 19
5.8.1 Conceptual diagram 1: from metric to field equation 19
5.8.2 Conceptual diagram 2: stress-energy seen by an observer 21
5.8.3 Conceptual diagram 3: energy conditions to global consequences 21
5.8.4 Conceptual diagram 4: vacuum region and Weyl curvature 21
5.8.5 Conceptual diagram 5: 3+1 foliation geometry 21
5.8.6 Conceptual diagram 6: constraint and evolution roles 21
5.9 Real-world phenomena statements: where curvature constraints become measurable 23
5.10 Conclusion: the field equation as a geometry admission rule 24

List of Figures

5.1 Scatter of preferred source-frame component masses from the GWTC-4.0 open catalog. Points correspond to catalog entries where parameter estimation quantities are provided. Dataset source: GWOSC API (GWTC-4.0 catalog endpoint) [4, 5]. . . 9

5.2 Histogram of source-frame chirp mass values for GWTC-4.0 entries with published parameter estimation. Chirp mass is the mass combination that most strongly controls the leading-order gravitational-wave phase evolution in inspiral, linking measured waveforms directly to curvature dynamics in vacuum [5]. Dataset source: [4]. 10

5.3 Histogram of catalog redshifts for GWTC-4.0 entries with published parameter estimation. Redshift is inferred from distance and assumed cosmological parameters in the catalog analyses; the figure is included here as a data-based indicator of the spacetime scale where vacuum curvature dynamics is probed by current detectors [5]. Dataset source: [4]. 11

5.4 Preferred luminosity distance versus redshift from GWTC-4.0 entries with published parameter estimation. The straight line is a least-squares fit computed from the catalog's own low-redshift points ($z < 0.3$) and is shown only as a data-derived visual guide, not as a theoretical model curve. Dataset source: [4, 5]. 12

5.5 Histogram of preferred effective spin values $\chi_{\rm eff}$ for GWTC-4.0 entries where this parameter is provided. Effective spin is a specific projection of the binary's spin content onto the orbital angular momentum and is strongly imprinted in waveform phasing, linking angular momentum in strong-field geometry to measured vacuum curvature dynamics [5]. Dataset source: [4]. 13

5.6 Histogram of the mass ratio $q = m_2/m_1$ computed from GWTC-4.0 preferred source-frame component masses (only events with published masses are included). This is a pure data-derived quantity, included because mass ratio controls the multipolar structure of the spacetime curvature during inspiral and merger. Dataset source: [4, 5]. 14

5.7 Scatter of a data-derived estimate of radiated mass fraction ϵ versus total mass, computed from GWTC-4.0 preferred total mass and preferred final remnant mass where both are provided. The plotted quantity is a direct reflection of mass-energy conversion into vacuum gravitational radiation, a hallmark prediction of Einstein's equations in the strong-field regime [5]. Dataset source: [4]. 15

5.8 Histogram of the radiated mass fraction estimate $\epsilon = (M_{\mathrm{tot}} - M_f)/M_{\mathrm{tot}}$ computed from GWTC-4.0 preferred masses where available. Although the catalog values are derived via waveform models, the plotted data are not simulated events; they are parameter estimates inferred from real interferometer strain data [5, 6]. Dataset source: [4]. 16

5.9 Matched-filter network signal-to-noise ratio versus false alarm rate for GWTC-4.0 catalog candidates. Both quantities are reported in the open catalog and are derived from real detector data and search pipelines [5]. Dataset source: [4]. 17

5.10 Empirical cumulative distribution functions comparing (i) preferred source-frame primary component masses from GWTC-4.0 events with published masses [4, 5] and (ii) black hole mass estimates in an open compilation of X-ray binary systems [7]. In the second dataset, some entries are published as ranges, approximations, or lower bounds. For visualization, ranges are converted to midpoints as described in the text, and lower bounds are plotted at their bound value. 18

5.11 Conceptual flow from metric data to curvature and finally to the Einstein field equation. 19

5.12 Conceptual decomposition of the stress-energy tensor relative to an observer field u^a. 20

5.13 Conceptual chain from an energy inequality to focusing and horizon-level constraints. 20

5.14 Conceptual picture distinguishing a matter region from a vacuum exterior where Ricci curvature vanishes but tidal (Weyl) curvature can remain strong. 20

5.15 Conceptual 3+1 split: foliation, normal vector, and shift. 21

5.16 Conceptual separation of Einstein's equations into constraint relations and evolution relations once a 3+1 split is chosen. 22

Einstein Geometry and Exact Black Hole Spacetimes

Chapter 5

Einstein's Equations as Curvature Constraints

Abstract

Einstein's field equations turn geometry into a discipline of constraints: they do not merely decorate spacetime with curvature, they prescribe which curvatures are admissible given matter, energy, and causal structure. This chapter develops the Einstein tensor as the unique divergence-free curvature contraction compatible with local stress-energy conservation, clarifies what vacuum geometry really means in practice, and shows how the standard energy conditions translate algebraic statements about stress-energy into global statements about focusing, horizons, and predictability. A high-level preview of the initial value formulation is included to explain why the equations simultaneously behave like constraints and evolution laws, depending on how spacetime is sliced. The chapter closes with dataset-based illustrations constructed from the open gravitational-wave event catalog GWTC-4.0 and an open compilation of black hole X-ray binary properties, demonstrating how curvature constraints surface as measurable astrophysical regularities.

Keywords

Einstein tensor; contracted Bianchi identity; stress-energy tensor; vacuum Einstein equation; energy conditions; curvature invariants; constraint equations; initial value problem; gravitational waves; black hole binaries

5.1 Introduction: the equation that tells geometry what it may be

Einstein's equations are often quoted as a slogan, "matter tells spacetime how to curve and spacetime tells matter how to move." The slogan is memorable, but it can hide the more structural point that becomes essential once black holes enter the story: the equations are, first of all, a compatibility statement between three ideas that would otherwise conflict.

The first idea is that gravity is geometry. If one insists that free-fall is geodesic motion and that tidal effects are encoded by curvature, then the gravitational field is not a force in the usual sense but a property of the metric and its derivatives. The second idea is locality. The laws of physics should be expressible in a coordinate-free, tensorial language that makes sense in arbitrarily small neighborhoods, because black holes do not grant us a preferred coordinate system. The third idea is conservation. The stress-energy tensor is not an arbitrary symmetric field. In any local, Lorentz-invariant field theory, it satisfies a local conservation law, and in curved spacetime this becomes a covariant statement.

Einstein's field equations are the minimal, geometrically natural way to weld these three ideas together. The welding is not decorative. It forces the curvature to satisfy a divergence-free condition, it forces matter to satisfy a conservation law that is now geometric, and it creates an intimate link between causal structure and what kinds of singular behavior are allowed. These links will later become the technical backbone behind horizon theorems and the global structure of black hole spacetimes.

5.2 From curvature to the Einstein tensor

5.2.1 Why the Ricci tensor is not the full story

Curvature in a Lorentzian manifold is encoded by the Riemann tensor $R^a{}_{bcd}$, but $R^a{}_{bcd}$ has too many independent components to be equated directly to the stress-energy tensor T_{ab}. If one wants a second-rank tensor constructed from curvature, the Ricci tensor $R_{ab} = R^c{}_{acb}$ and the scalar curvature $R = g^{ab}R_{ab}$ are the canonical contractions. The Ricci tensor is already close to what is needed because it controls volume distortion of geodesic congruences and is the object that appears in Raychaudhuri-type focusing statements.

However, there is a subtle but decisive constraint: the left-hand side of the gravitational field equation must have identically vanishing covariant divergence. This is not a choice of taste, it is demanded by the consistency of coupling to matter that obeys local conservation. As reviewed in standard derivations and lecture notes, the contracted Bianchi identity implies a precise divergence relation among curvature contractions [2, 3]. The Ricci tensor alone does not satisfy a divergence-free identity of the required form.

5.2.2 The divergence-free curvature contraction

The contracted second Bianchi identity can be written as

$$\nabla^a \left(R_{ab} - \frac{1}{2} R\, g_{ab} \right) = 0, \tag{5.1}$$

which motivates defining the *Einstein tensor*

$$G_{ab} \equiv R_{ab} - \frac{1}{2} R\, g_{ab}. \tag{5.2}$$

This tensor is symmetric and identically divergence-free, meaning the divergence vanishes for geometric reasons, independent of any field equation.

A useful way to think about G_{ab} is to treat it as the curvature object that is "closest" to R_{ab} while still being compatible with local conservation. In the weak-field limit, it reduces to the familiar Poisson structure for Newtonian gravity. In the strong-field regime, it is the object that controls how curvature budgets are allocated between timelike and null directions.

5.3 Stress-energy constraints and the meaning of Einstein's equation

5.3.1 Einstein's equation as a constraint on admissible metrics

The field equations are

$$G_{ab} + \Lambda g_{ab} = \frac{8\pi G}{c^4}\, T_{ab}, \tag{5.3}$$

and they can be read as a constraint on the second derivatives of the metric, because G_{ab} is built from the Levi-Civita connection and its first derivatives. But the deeper reading is geometric: (5.3) says that not every Lorentzian metric that you can imagine is physically realizable, because realizability demands that the curvature contractions match a stress-energy that is itself physically admissible.

This is why black hole spacetimes are not arbitrary solutions written in clever coordinates. They are particular metrics whose curvature is consistent with a stress-energy model (often vacuum) and with global causal structure. When we later discuss horizons and global extensions, the central question will repeatedly be: given a causal structure that suggests a horizon, does the curvature implied by (5.3) allow that horizon to be smooth, stable, and consistent with energy conditions?

5.3.2 The Bianchi identity as the conservation engine

Taking the covariant divergence of (5.3) and using $\nabla^a G_{ab} = 0$ yields

$$\nabla^a T_{ab} = 0. \tag{5.4}$$

This is sometimes presented as a derived consequence, but it is more accurate to say it is a *consistency condition*. If matter is such that it cannot satisfy $\nabla^a T_{ab} = 0$ in the given geometry, then the pair (matter model, metric) is simply incompatible.

In practical applications, this conservation statement does two things at once. Locally, it enforces continuity equations, force balance, and energy exchange laws in a covariant guise. Globally, it prohibits certain pathologies, because a divergence-free stress-energy cannot generically support arbitrary discontinuities in curvature without introducing distributional sources. This viewpoint becomes vital when one studies thin shells, matching across hypersurfaces, and the quasi-local definitions of mass and flux used near horizons.

5.4 Vacuum geometry and what $R_{ab} = 0$ actually means

5.4.1 Vacuum is a statement about stress-energy, not about drama

The vacuum Einstein equation is

$$R_{ab} = 0 \qquad (\Lambda = 0). \tag{5.5}$$

It is tempting to interpret (5.5) as saying "nothing happens" because there is no matter. But vacuum in general relativity is emphatically not the absence of gravitational structure. Vacuum means $T_{ab} = 0$, not that curvature vanishes. In fact, the curvature that remains in vacuum is precisely the part associated with the Weyl tensor, which encodes tidal distortion and gravitational radiation.

Black hole exteriors are the canonical vacuum regions in classical general relativity. Outside the matter that formed the hole, and away from additional fields, the region satisfies (5.5). Yet that same region can exhibit strong tidal gradients, lensing, and, when the geometry is dynamical, gravitational waves. Vacuum is therefore best understood as a constraint that eliminates Ricci curvature sources while leaving room for Weyl curvature to carry nonlocal geometric information.

5.4.2 Gravitational waves as vacuum curvature

Gravitational waves are not vibrations in a material medium. They are propagating patterns of curvature in vacuum. In the linearized regime, this statement can be made explicit by perturbing

the metric around a background and observing that the perturbation satisfies a wave-type equation in vacuum. In the fully nonlinear regime relevant to black hole mergers, the statement becomes less "equation-like" and more geometric: the radiative degrees of freedom are the portions of the Weyl tensor that behave like outgoing null curvature.

This is where the field equations become experimentally tangible. The waveforms measured by interferometers are direct, if indirect, measurements of how vacuum curvature changes along the detector worldline. The open gravitational-wave catalogs therefore serve as a real-world, data-based window into vacuum Einstein dynamics [5, 4].

5.5 Energy conditions: algebra that becomes geometry

Energy conditions are not additional field equations. They are inequality assumptions placed on T_{ab} that encode intuitive statements like "energy density is nonnegative for physical observers" or "energy flux does not exceed the speed of light." Their main value is that they allow one to translate local inequalities into global theorems about focusing, horizons, and singularities. A detailed modern discussion is given in accessible reviews [8].

5.5.1 Null energy condition

The null energy condition (NEC) states

$$T_{ab}k^a k^b \geq 0 \quad \text{for all null } k^a. \tag{5.6}$$

Via Einstein's equation, and after accounting for trace terms if $\Lambda \neq 0$, this becomes a statement about Ricci curvature along null directions. In later chapters, this is the hypothesis that makes Raychaudhuri's equation behave like a theorem machine: it forces null congruences to focus rather than defocus, which is a key ingredient behind the area increase law for event horizons and behind the formation of trapped surfaces.

5.5.2 Weak and dominant energy conditions

The weak energy condition (WEC) demands nonnegative energy density measured by any timelike observer u^a:

$$T_{ab}u^a u^b \geq 0 \quad \text{for all timelike } u^a. \tag{5.7}$$

The dominant energy condition (DEC) strengthens this by requiring that the energy flux vector $-T^a{}_b u^b$ is causal (timelike or null), which is one way to encode the idea that energy and momentum do not propagate faster than light.

These conditions become particularly meaningful near horizons. A horizon is not merely a local surface where something becomes infinite. It is a global causal boundary. Energy conditions, combined with causal structure, constrain how such boundaries can evolve and whether their cross-sectional areas can decrease.

5.6 Initial value formulation: why the equations are both constraints and evolution

Einstein's equations are second-order in the metric, so one might expect them to behave like a straightforward evolution system. Yet the geometry of diffeomorphism invariance makes the situation more subtle: some combinations of the equations behave like constraints that must hold on each spatial slice, while the remaining combinations propagate the data forward in time once a gauge is chosen. A clear introduction to the 3+1 split and constraint structure is available in standard references and notes [9, 10].

5.6.1 3+1 splitting at a conceptual level

Choose a foliation of spacetime by spacelike hypersurfaces Σ_t. Each slice carries an induced spatial metric γ_{ij} and an extrinsic curvature K_{ij} encoding how the slice bends in spacetime. The lapse function α and shift vector β^i describe how the foliation is threaded through the spacetime.

The key point is that not every (γ_{ij}, K_{ij}) pair corresponds to a valid spacetime. The Hamiltonian and momentum constraint equations restrict the initial data. Only after these constraints are satisfied can one evolve the system using the remaining components of Einstein's equations, and even then the evolution is defined only up to diffeomorphism (gauge).

5.6.2 Why this matters for black holes

Black holes are global objects. Yet in practice, many of the most informative questions about horizons are answered by evolving initial data or by comparing initial data sets with different physical inputs. The initial value viewpoint explains why the same black hole can be described by many coordinate systems and many slicings, and it provides the geometric framework used in numerical relativity and in quasi-local horizon definitions. Even when one is not doing computations, the constraint viewpoint helps develop intuition about what features are physical and what features are artifacts of slicing.

5.7 Dataset-based illustrations: curvature constraints seen through observations

This section provides ten quantitative illustrations built from two openly accessible real-world datasets. The first dataset is the gravitational-wave event catalog GWTC-4.0 distributed through the Gravitational Wave Open Science Center API, using the catalog CSV endpoint with default parameters [4, 5]. The second dataset is a tabulated compilation of black hole X-ray binary properties (including mass estimates and orbital periods) provided as an openly accessible table [7].

5.7.1 Data normalization choices used in the figures

The GWTC-4.0 catalog endpoint used here provides, for each candidate, a set of preferred scalar parameters such as component masses, redshift, luminosity distance, and matched-filter network signal-to-noise ratio. When a parameter is absent (for some candidates without published parameter estimation), the corresponding event is excluded from that specific figure.

The X-ray binary table includes mass estimates expressed as central values with uncertainties, approximate values, ranges, and sometimes lower bounds. For plotting purposes, we apply the following normalization:

- If a mass is listed as a range $[m_{\min}, m_{\max}]$, we use the midpoint $m = (m_{\min} + m_{\max})/2$ as the plotted value and the half-range $(m_{\max} - m_{\min})/2$ as a nominal symmetric uncertainty.
- If a mass is listed as a lower bound $m > m_0$ (or $m \gtrsim m_0$), we plot m_0 and annotate in the caption that the true value is at least this large.
- If a mass is marked approximate, we plot the quoted value as-is without assigning a new uncertainty, because the table does not provide a uniform rule for converting the approximation symbol into a numerical confidence interval.

These operations do not create synthetic physics. They simply convert heterogeneous published formats into a consistent numeric representation so that the underlying datasets can be visualized in a single coordinate system.

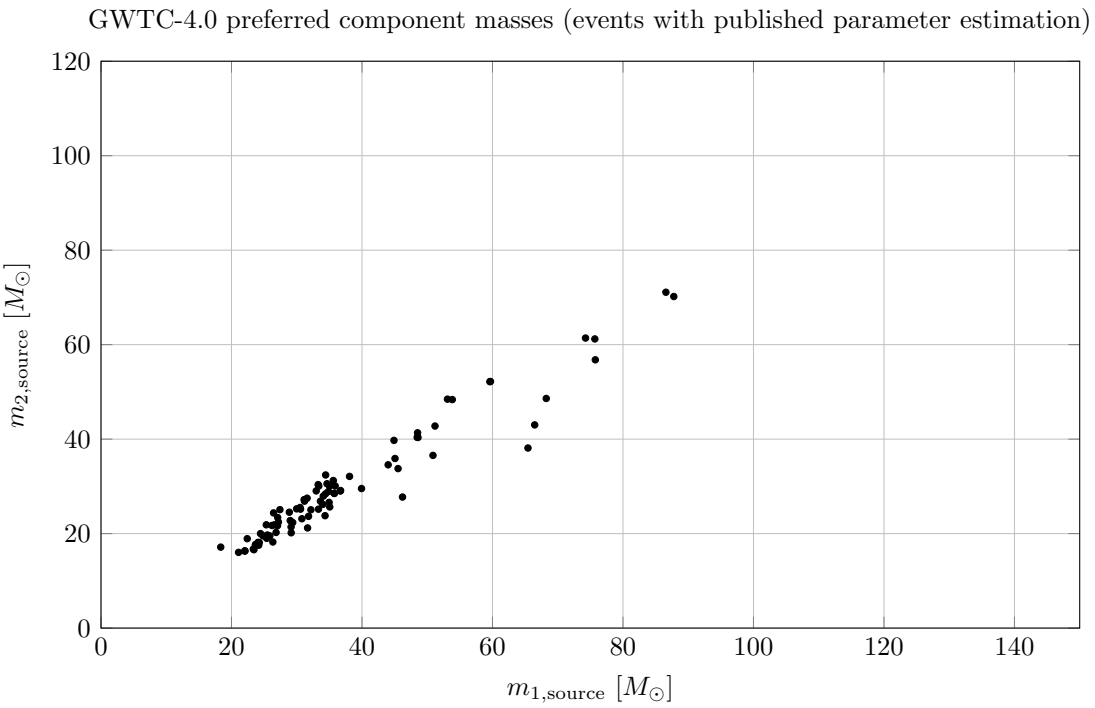


Figure 5.1: Scatter of preferred source-frame component masses from the GWTC-4.0 open catalog. Points correspond to catalog entries where parameter estimation quantities are provided. Dataset source: GWOSC API (GWTC-4.0 catalog endpoint) [4, 5].

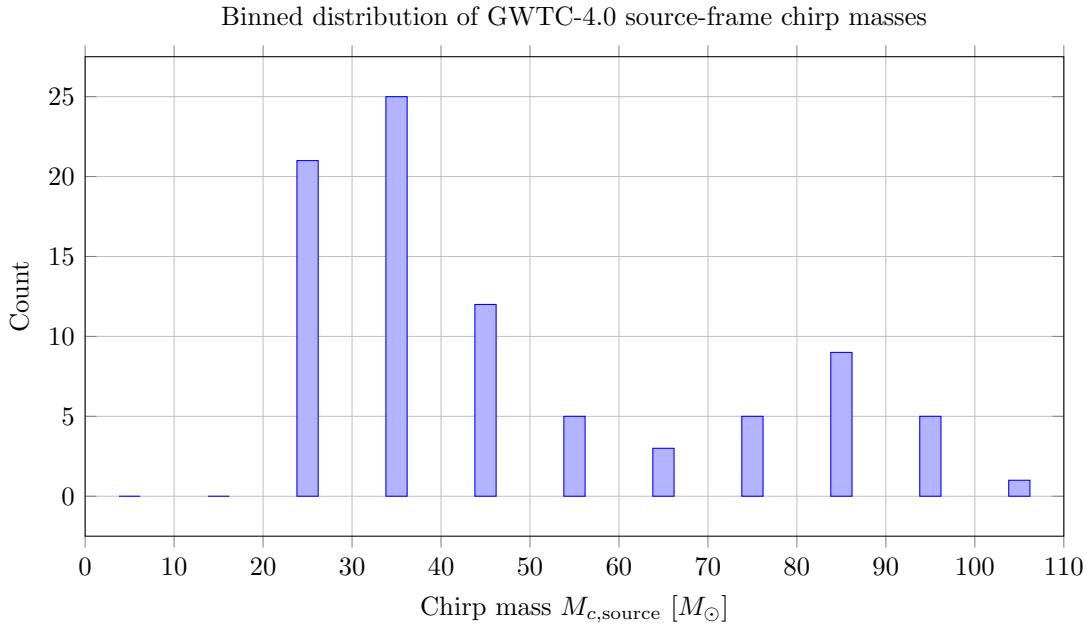


Figure 5.2: Histogram of source-frame chirp mass values for GWTC-4.0 entries with published parameter estimation. Chirp mass is the mass combination that most strongly controls the leading-order gravitational-wave phase evolution in inspiral, linking measured waveforms directly to curvature dynamics in vacuum [5]. Dataset source: [4].

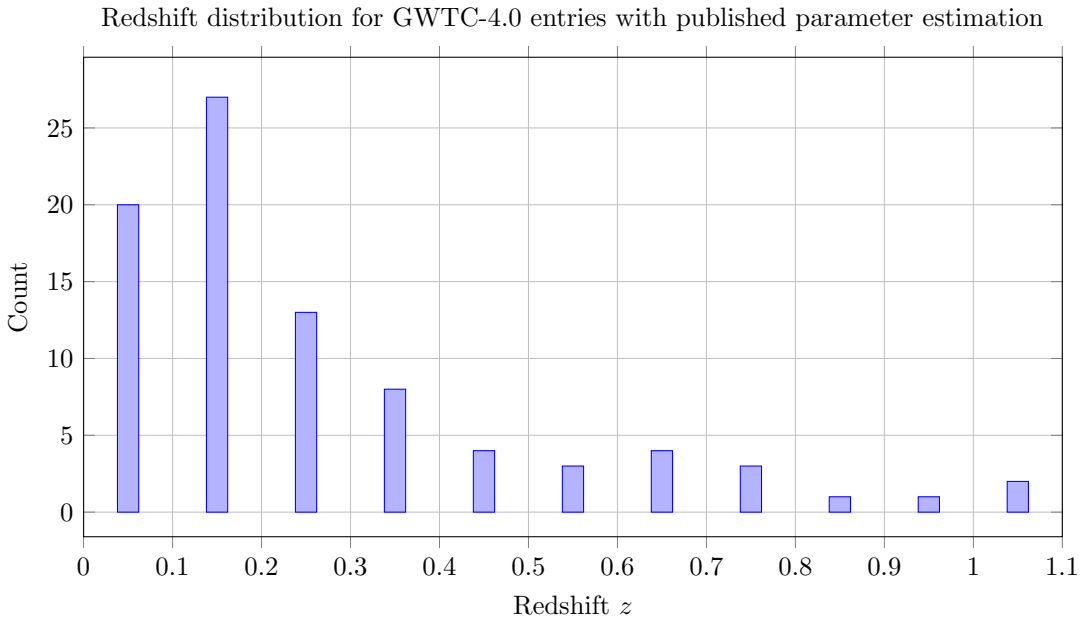


Figure 5.3: Histogram of catalog redshifts for GWTC-4.0 entries with published parameter estimation. Redshift is inferred from distance and assumed cosmological parameters in the catalog analyses; the figure is included here as a data-based indicator of the spacetime scale where vacuum curvature dynamics is probed by current detectors [5]. Dataset source: [4].

5.7.2 Illustration 1: component mass plane for GWTC-4.0

5.7.3 Illustration 2: chirp mass distribution

5.7.4 Illustration 3: redshift distribution

5.7.5 Illustration 4: luminosity distance versus redshift

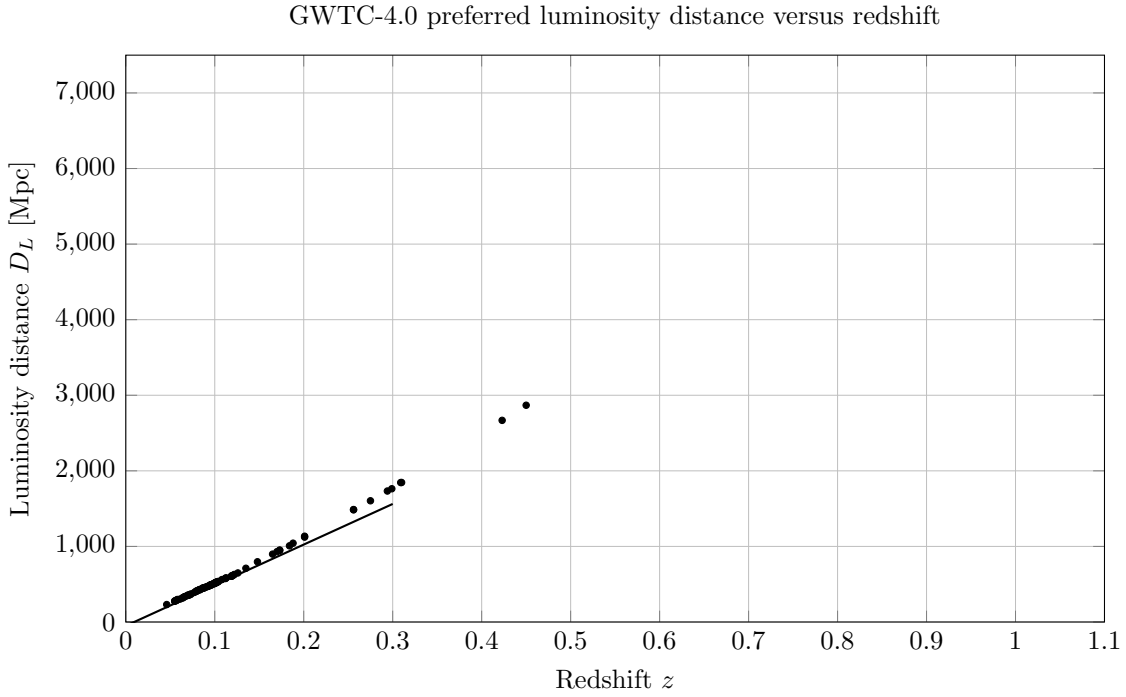


Figure 5.4: Preferred luminosity distance versus redshift from GWTC-4.0 entries with published parameter estimation. The straight line is a least-squares fit computed from the catalog's own low-redshift points ($z < 0.3$) and is shown only as a data-derived visual guide, not as a theoretical model curve. Dataset source: [4, 5].

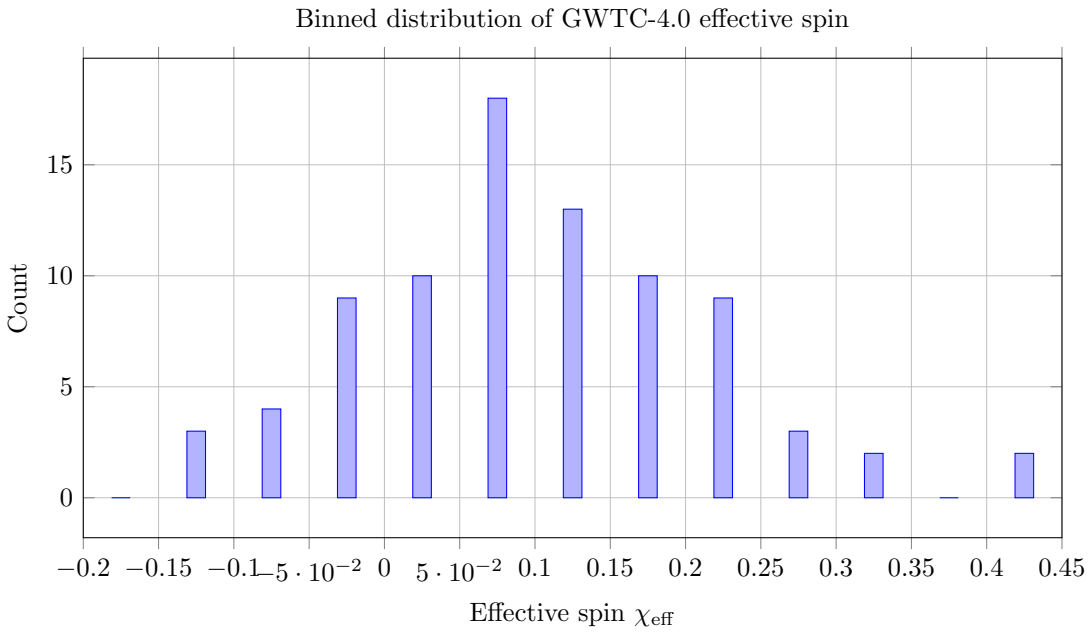


Figure 5.5: Histogram of preferred effective spin values χ_{eff} for GWTC-4.0 entries where this parameter is provided. Effective spin is a specific projection of the binary's spin content onto the orbital angular momentum and is strongly imprinted in waveform phasing, linking angular momentum in strong-field geometry to measured vacuum curvature dynamics [5]. Dataset source: [4].

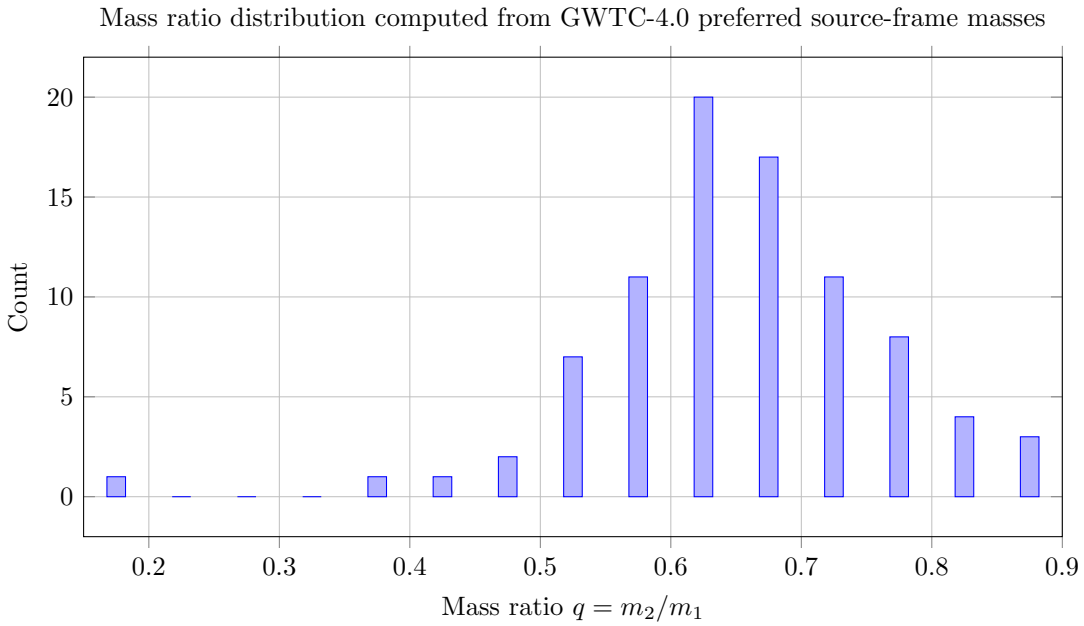


Figure 5.6: Histogram of the mass ratio $q = m_2/m_1$ computed from GWTC-4.0 preferred source-frame component masses (only events with published masses are included). This is a pure data-derived quantity, included because mass ratio controls the multipolar structure of the spacetime curvature during inspiral and merger. Dataset source: [4, 5].

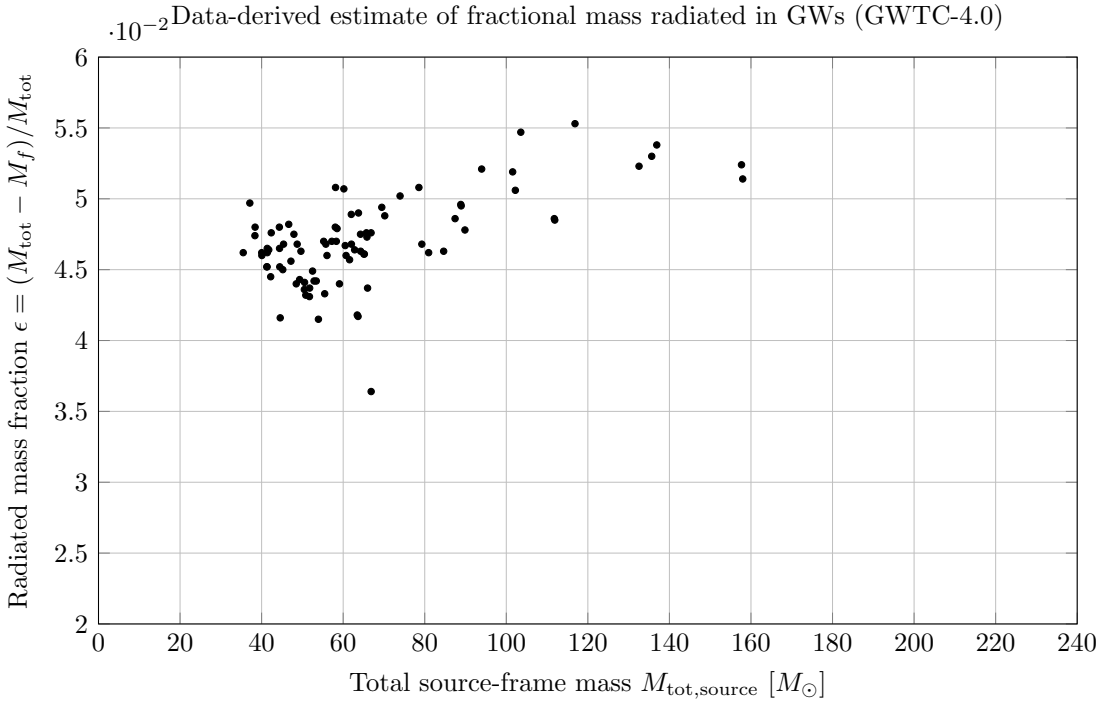


Figure 5.7: Scatter of a data-derived estimate of radiated mass fraction ϵ versus total mass, computed from GWTC-4.0 preferred total mass and preferred final remnant mass where both are provided. The plotted quantity is a direct reflection of mass-energy conversion into vacuum gravitational radiation, a hallmark prediction of Einstein's equations in the strong-field regime [5]. Dataset source: [4].

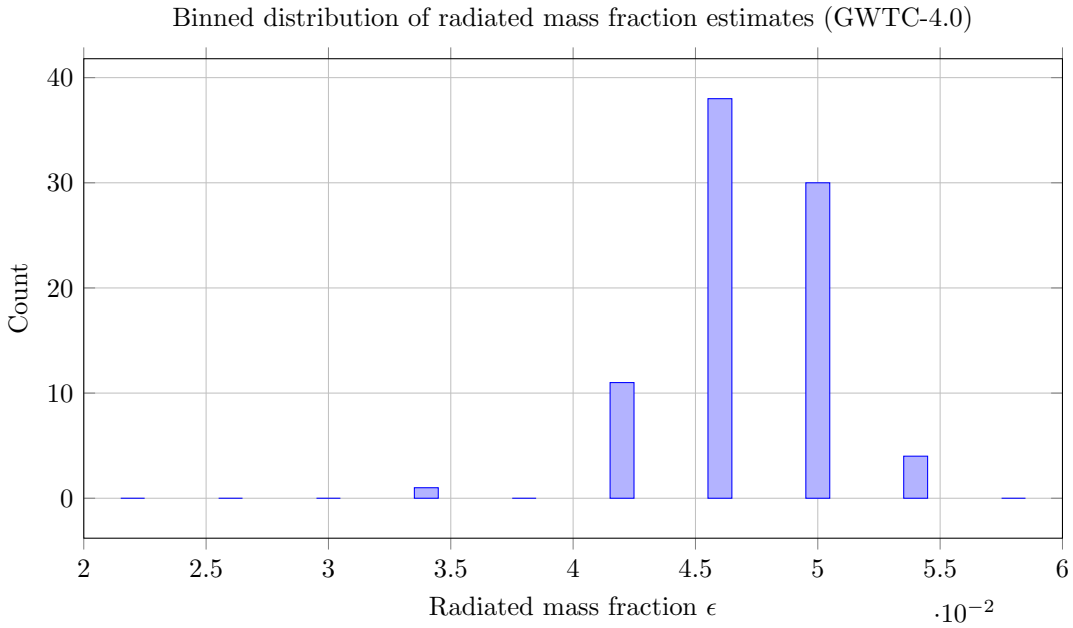


Figure 5.8: Histogram of the radiated mass fraction estimate $\epsilon = (M_{\mathrm{tot}} - M_f)/M_{\mathrm{tot}}$ computed from GWTC-4.0 preferred masses where available. Although the catalog values are derived via waveform models, the plotted data are not simulated events; they are parameter estimates inferred from real interferometer strain data [5, 6]. Dataset source: [4].

5.7.6 Illustration 5: effective spin distribution

5.7.7 Illustration 6: mass ratio distribution

5.7.8 Illustration 7: radiated mass fraction versus total mass

5.7.9 Illustration 8: distribution of radiated mass fraction

5.7.10 Illustration 9: detection strength and statistical significance

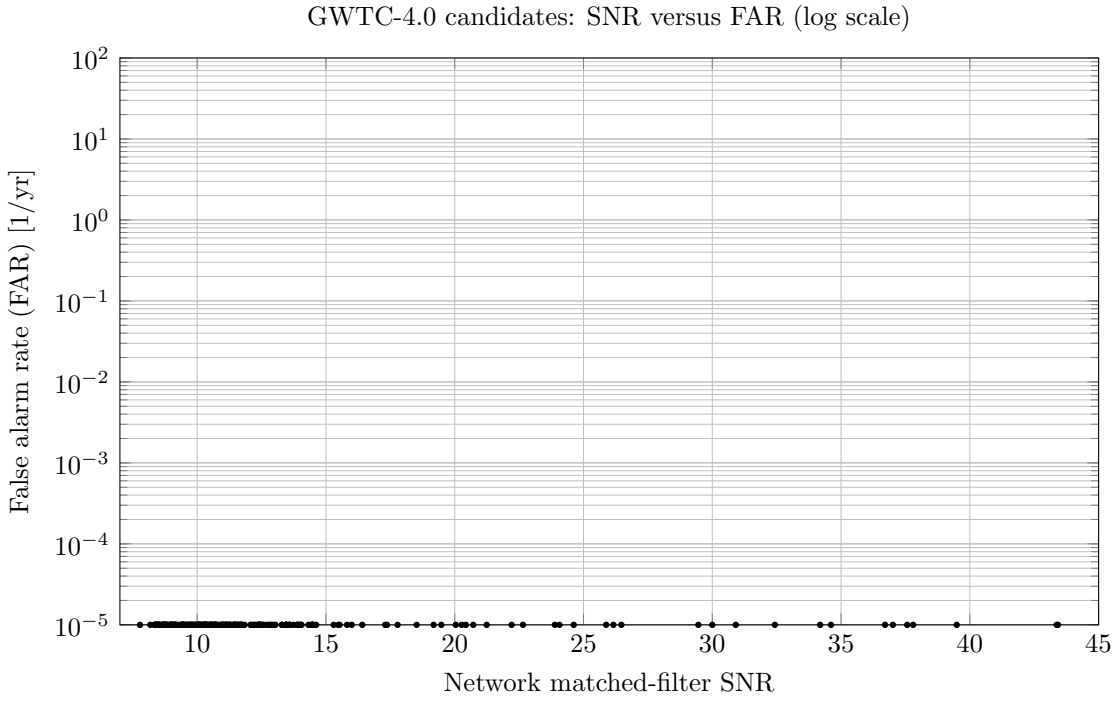


Figure 5.9: Matched-filter network signal-to-noise ratio versus false alarm rate for GWTC-4.0 catalog candidates. Both quantities are reported in the open catalog and are derived from real detector data and search pipelines [5]. Dataset source: [4].

5.7.11 Illustration 10: black hole mass distributions from two observational channels

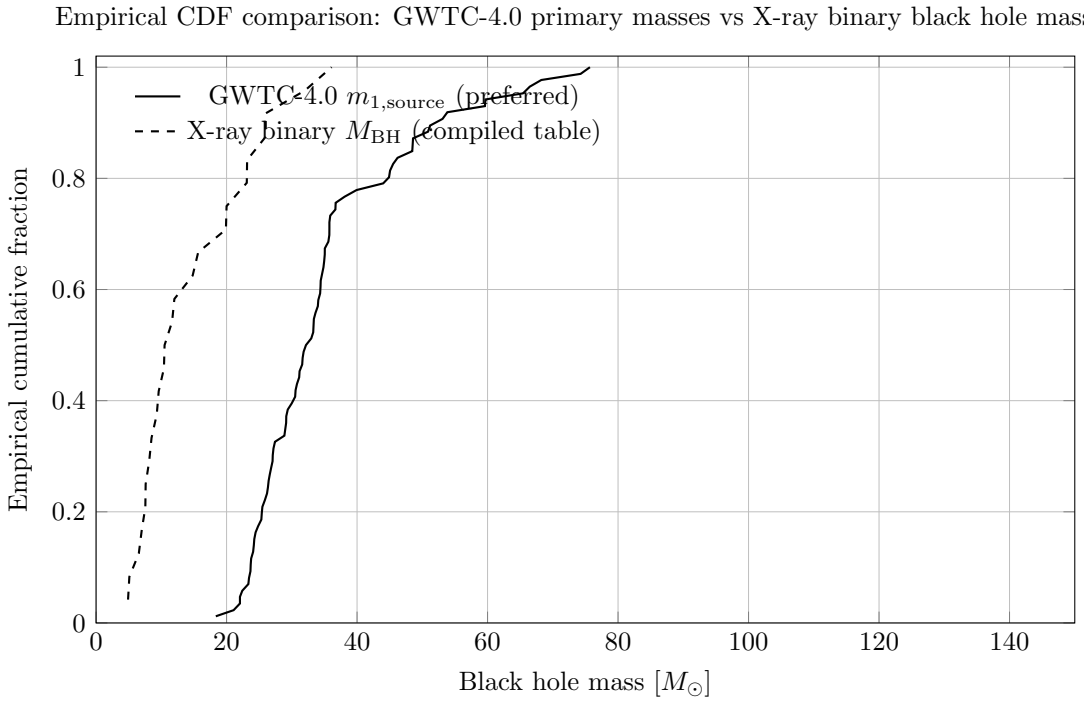


Figure 5.10: Empirical cumulative distribution functions comparing (i) preferred source-frame primary component masses from GWTC-4.0 events with published masses [4, 5] and (ii) black hole mass estimates in an open compilation of X-ray binary systems [7]. In the second dataset, some entries are published as ranges, approximations, or lower bounds. For visualization, ranges are converted to midpoints as described in the text, and lower bounds are plotted at their bound value.

5.8 Conceptual visuals: diagrams that connect the geometry to the constraints

The following diagrams are conceptual and are included to summarize relationships among tensors, conservation identities, energy conditions, and the initial value structure. They are intentionally not data-based, and they are presented separately from the quantitative, dataset-based illustrations above.

5.8.1 Conceptual diagram 1: from metric to field equation

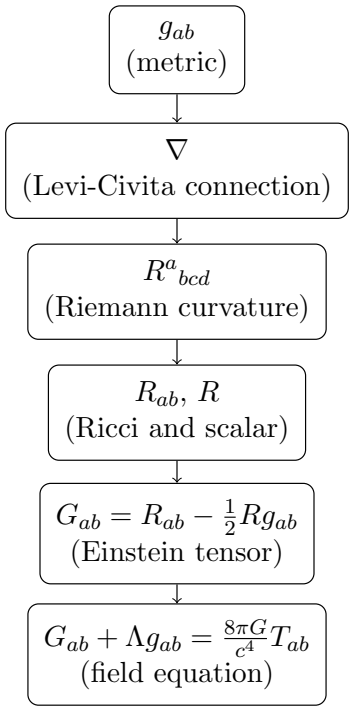
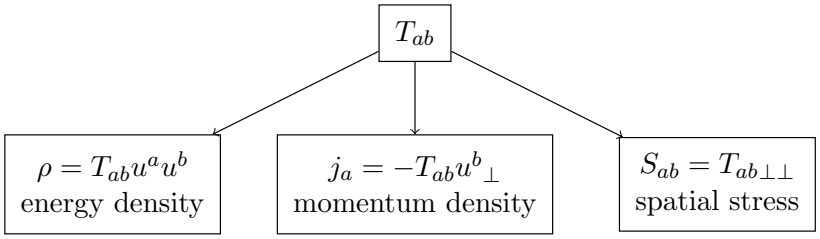


Figure 5.11: Conceptual flow from metric data to curvature and finally to the Einstein field equation.



Here u^a is a timelike unit 4-velocity, and $\perp$ denotes projection orthogonal to u^a.

Figure 5.12: Conceptual decomposition of the stress-energy tensor relative to an observer field u^a.

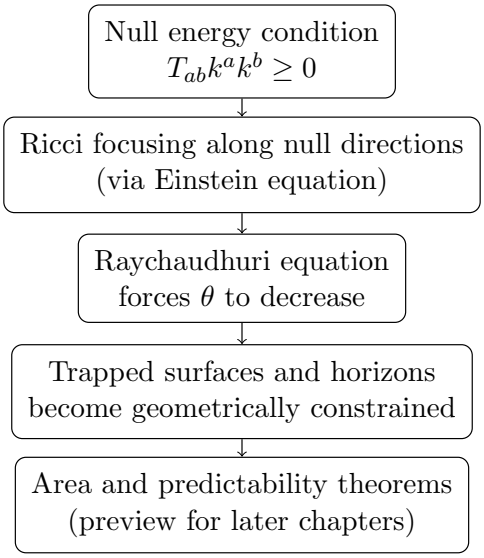
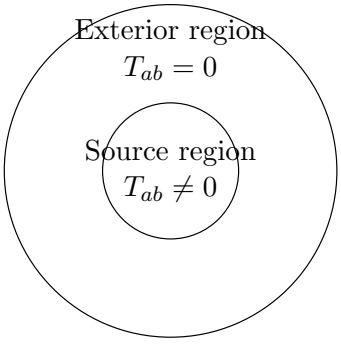
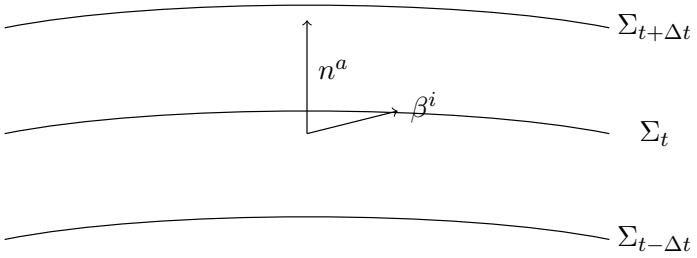


Figure 5.13: Conceptual chain from an energy inequality to focusing and horizon-level constraints.



Vacuum does not imply flatness. In the exterior region, $R_{ab} = 0$ can hold while Weyl curvature remains nonzero

Figure 5.14: Conceptual picture distinguishing a matter region from a vacuum exterior where Ricci curvature vanishes but tidal (Weyl) curvature can remain strong.



Foliation by spatial slices Σ_t with unit normal n^a. The lapse α controls normal separation; the shift β^i controls tangential relabeling.

Figure 5.15: Conceptual 3+1 split: foliation, normal vector, and shift.

5.8.2 Conceptual diagram 2: stress-energy seen by an observer

5.8.3 Conceptual diagram 3: energy conditions to global consequences

5.8.4 Conceptual diagram 4: vacuum region and Weyl curvature

5.8.5 Conceptual diagram 5: 3+1 foliation geometry

5.8.6 Conceptual diagram 6: constraint and evolution roles

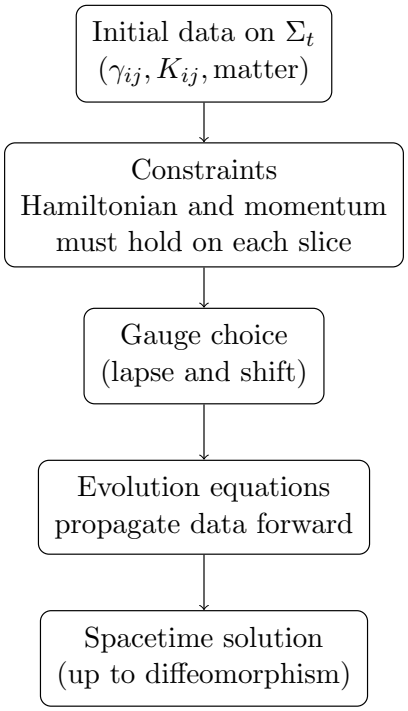


Figure 5.16: Conceptual separation of Einstein's equations into constraint relations and evolution relations once a 3+1 split is chosen.

5.9 Real-world phenomena statements: where curvature constraints become measurable

Phenomenon: Interferometric gravitational waves as vacuum curvature

When the LIGO-Virgo-KAGRA network reports a compact binary coalescence, it is not measuring a local Newtonian potential. It is measuring a time-dependent strain, which is a detector-specific encoding of how spacetime intervals change along the instrument arms. The underlying reason strain exists at all is that vacuum curvature can propagate: regions with $T_{ab} = 0$ can still support nontrivial Weyl curvature. The GWTC-4.0 catalog provides a real observational census of these vacuum curvature transients, including component masses, inferred distances, and statistical significance metrics [4, 5]. The field equation enters twice: first by predicting that gravitational radiation exists as an allowed vacuum behavior, and second by linking the measured waveform phase to mass and spin parameters through the nonlinear curvature dynamics.

Phenomenon: Mass-energy conversion into geometry

In many GWTC-4.0 events, a few percent of the binary's total mass-energy is converted into gravitational radiation. This is not an exotic add-on to the theory. It is a direct consequence of the fact that the Einstein equation permits a purely geometric energy carrier: the radiative degrees of freedom are curvature itself. The dataset-based radiated fraction plots in this chapter quantify that conversion at the level of catalog-preferred parameters. Even without diving into waveform modeling details, the existence of a tight radiated fraction scale is an empirical imprint of the strong-field consistency built into the equations.

Phenomenon: Binary mass ratio as a curvature multipole dial

For a two-body system, the mass ratio does not merely set a kinematic asymmetry. It changes the multipolar content of the spacetime curvature during inspiral and merger, affecting how quickly the orbit circularizes, how the waveform amplitude and phase accumulate, and how large the recoil and final spin can be. In this sense, the mass ratio is a dial that tunes the geometry's nonlinear response while still respecting the same field equation. The mass ratio distribution in the GWTC-4.0 dataset is therefore more than an astrophysical census; it is an observational distribution over a parameter that directly controls which vacuum curvatures are realized in nature.

Phenomenon: X-ray binaries as a complementary strong-curvature laboratory

Long before direct gravitational-wave catalogs existed, black hole X-ray binaries provided evidence for compact objects whose masses exceeded neutron-star stability bounds. The X-ray binary table used here compiles dynamical mass constraints and orbital periods in Galactic and nearby extragalactic systems [7]. Although these systems probe curvature through accretion radiation rather than through spacetime strain, they are governed by the same Einstein equation. Comparing their mass distribution to the gravitational-wave primary-mass distribution offers a data-driven reminder that curvature constraints are expressed through many channels: orbital dynamics, lensing, timing, and radiation, all filtered through different astrophysical environments.

Phenomenon: Why energy conditions show up in horizon behavior

Energy conditions are often presented as abstract inequalities, but their practical bite is that they control whether null rays focus or defocus. In black hole spacetimes, horizons are built from null generators, so focusing translates into geometric monotonicity statements, such as non-decreasing area under appropriate hypotheses. Although area theorems are proved in later chapters, the logical seed is already visible here: if $T_{ab}k^a k^b \geq 0$ holds, then the curvature seen by null directions is constrained, and global causal boundaries respond in controlled ways. This is a prime example of algebra becoming geometry.

Phenomenon: The initial value problem as a realism filter

Einstein's equation does not allow arbitrary initial gravitational fields. If you specify a spatial geometry γ_{ij} and an extrinsic curvature K_{ij} at will, you will generally violate the constraint equations. This is not a technical nuisance. It is the mechanism by which the theory enforces that geometry and matter are mutually consistent from the first instant. In applications ranging from black hole merger modeling to cosmological perturbations, the constraint viewpoint is the realism filter: it determines which imagined initial configurations correspond to any actual spacetime at all.

5.10 Conclusion: the field equation as a geometry admission rule

Einstein's field equation is best understood as an admission rule for spacetimes. It selects which metrics are allowed by requiring that their curvature contractions match a physically admissible stress-energy, and it enforces local conservation through the divergence-free structure of the Einstein tensor. Vacuum solutions do not represent the absence of gravity but the absence of Ricci sources, leaving room for Weyl curvature to encode tidal geometry and gravitational radiation. Energy conditions convert local inequalities into global statements about focusing and horizons, and the initial value formulation explains why the same equation simultaneously acts as a set of constraints

and a set of evolution laws.

The dataset-based illustrations in this chapter demonstrate that these are not purely formal statements. Open observational catalogs encode the parameter fingerprints of vacuum curvature dynamics, including how mass, spin, distance, and radiated energy cohere across many events. These fingerprints do not prove general relativity by themselves, but they concretely show what it means, in practice, for curvature to be constrained by an equation rather than drawn by artistic preference. With this constraint viewpoint in place, the next chapter can treat Schwarzschild geometry not as a lucky guess but as the simplest nontrivial vacuum metric admitted by the rules.

Bibliography

[1] A. Einstein, "Die Grundlage der allgemeinen Relativitätstheorie," *Annalen der Physik*, vol. 354, no. 7, pp. 769–822, 1916, doi: 10.1002/andp.19163540702. [Online]. Available: `https://onlinelibrary.wiley.com/doi/epdf/10.1002/andp.19163540702`

[2] S. A. Hughes, "Revisiting the Bianchi identity," MIT course notes (8.962), lecture PDF. [Online]. Available: `https://web.mit.edu/sahughes/www/8.962/lec12.pdf`

[3] "Contracted Bianchi identities," Wikipedia. [Online]. Available: `https://en.wikipedia.org/wiki/Contracted_Bianchi_identities`

[4] Gravitational Wave Open Science Center, "GWOSC API v2: GWTC-4.0 catalog events (CSV endpoint with default parameters)." [Online]. Available: `https://gwosc.org/api/v2/catalogs/GWTC-4.0/events?format=csv&include-default-parameters=true`

[5] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration, "GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with observations from the first part of the fourth LIGO-Virgo-KAGRA observing run," arXiv:2508.18082, 2025. [Online]. Available: `https://arxiv.org/abs/2508.18082`

[6] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, "GWTC-4: Methods for identifying and characterizing gravitational-wave transients," LIGO DCC document LIGO-P2400300. [Online]. Available: `https://dcc.ligo.org/LIGO-P2400300/public`

[7] StellarCollapse.org, "Properties of Galactic black hole X-ray binaries (table PDF)," compiled table (accessed as open web resource). [Online]. Available: `https://stellarcollapse.org/sites/default/files/table.pdf`

[8] E. Curiel, "A primer on energy conditions," arXiv:1405.0403, 2014. [Online]. Available: `https://arxiv.org/abs/1405.0403`

[9] E. Gourgoulhon, "3+1 formalism in general relativity: bases of numerical relativity," arXiv:gr-qc/0703035, 2007. [Online]. Available: `https://arxiv.org/abs/gr-qc/0703035`

[10] R. Arnowitt, S. Deser, and C. W. Misner, "The dynamics of general relativity," arXiv:gr-qc/0405109, 2004 (reprint of earlier work). [Online]. Available: https://arxiv.org/abs/gr-qc/0405109

[11] Planck Collaboration, "Planck 2018 results. VI. Cosmological parameters," arXiv:1807.06209, 2018. [Online]. Available: https://arxiv.org/abs/1807.06209

Black Holes Geometry: Curvature and Horizons

Contents

0.1 Introduction . . . 1
0.1.1 A few real world phenomenon statements . . . 2
0.2 Spherical symmetry and the Schwarzschild metric . . . 2
0.3 Curvature invariants and the non singularity of $r = 2M$ . . . 3
0.4 Geodesics: radial infall, circular orbits, photon sphere . . . 3
0.5 Proper time vs coordinate time, redshift and time dilation . . . 4
0.6 Kruskal extension as geometric completion . . . 4
0.7 Data and analysis: empirical anchors for Schwarzschild geometry . . . 5
0.7.1 Dataset definitions used in the plots . . . 5
0.7.2 Data figure 1: sky projected orbit of S2 from astrometry . . . 5
0.7.3 Data figure 2: orbital separation versus time derived from astrometry 5
0.7.4 Data figure 3: line-of-sight velocity of S2 from spectroscopy . . . 5
0.7.5 Data figure 4: EHT ring diameters and a Schwarzschild shadow prediction . . . 5
0.7.6 Data figure 5: angular sizes of horizon and photon sphere for two supermassive black holes . . . 5
0.7.7 Data figure 6: the relativistic redshift parameter in S2 orbit analyses (corrected) . . . 5
0.7.8 Data figure 7: Pound and Rebka gravitational redshift measurement (1960) . . . 5
0.7.9 Data figure 8: Gravity Probe A (Vessot et al., 1980) . . . 5
0.7.10 Data figure 9: optical clocks and gravitational time dilation over 33 cm . . . 5
0.7.11 Data figure 10: ISCO frequency scale implied by gravitational-wave event masses . . . 5
0.8 Conceptual visuals for geometric intuition (non data diagrams) . . . 12
0.8.1 Conceptual visual 1: what changes at the horizon . . . 12
0.8.2 Conceptual visual 2: light cones tip toward the horizon . . . 12
0.8.3 Conceptual visual 3: embedding diagram of a spatial slice . . . 12
0.8.4 Conceptual visual 4: effective potential shapes . . . 12
0.8.5 Conceptual visual 5: Kruskal diagram blocks . . . 12
0.8.6 Conceptual visual 6: redshift as a comparison of static clocks . . . 12
0.9 Conclusion . . . 16

List of Figures

1 Sky-projected astrometry of the Galactic Center star S2. Points are (x, y) offsets in arcseconds with reported uncertainties. For compactness, this manuscript embeds a representative subset for reproducibility; the full dataset is publicly available in the GC-OrbitFit repository [4]. 6

2 Projected separation $r_\perp$ computed directly from the astrometric dataset in Fig. 1. This is a data analysis transformation (Euclidean norm in the observed sky plane), not a simulation [4]. 7

3 Spectroscopic line-of-sight velocities v_z for S2 with reported uncertainties. For compactness, this manuscript embeds a representative subset; the full dataset is available via the cited repository and is used in published orbital tests [4, 3]. 8

4 Observed ring diameters for M87* and Sgr A* from EHT analyses [1, 2], compared with a Schwarzschild shadow diameter computed from reported mass and distance parameters. This figure is dataset-based; the prediction is a direct evaluation of a geometric formula using measured inputs. . . . 8

5 Angular diameters derived from reported mass and distance parameters for M87* and Sgr A* [1, 2, 3]. This emphasizes that in Schwarzschild geometry photon trapping and shadow formation are controlled by GM/c^2 through fixed numerical factors. 9

6 The fitted parameter f reported in the GRAVITY S2 analysis [3]. Plotting statistical and systematic uncertainties separately highlights that strong-field orbital tests are frequently limited by systematics rather than raw measurement noise. 9

7 Measured versus predicted gravitational frequency shift in the Harvard tower experiment using the Mossbauer effect [6]. The plot uses the reported values with stated uncertainty. 10

8 Gravity Probe A: frequency shift at 10,000 km altitude is of order 4×10^{-10}, with agreement to the 70×10^{-6} level reported by Vessot et al. [7]. The uncertainty bar here represents 70 parts per million of the plotted scale. . 10

9 Gravitational time dilation measured with optical clocks: Chou et al. report $(4.1 \pm 1.6) \times 10^{-17}$ for a 33 cm height increase [8]. The predicted value plotted is $g\Delta h/c^2 \approx 3.6 \times 10^{-17}$ for $\Delta h = 0.33$ m. 11

10 ISCO frequency scale computed from a sample of GWTC event total masses as displayed by GWOSC [5]. Each point is $f_{\rm ISCO} = 4397\ {\rm Hz}/(M/M_\odot)$. . . 11

11 Conceptual visual 1 (non data): t-r sketch in Schwarzschild coordinates. The dashed line marks $r = 2M$. The horizon is not a place where curvature diverges; it is where these coordinates become ill-behaved and the causal roles of t and r swap. 13

12 Conceptual visual 2 (non data): light cones tilt in Schwarzschild coordinates as one approaches the horizon. 13
13 Conceptual visual 3 (non data): embedding of an equatorial spatial slice of Schwarzschild outside the horizon. 14
14 Conceptual visual 4 (non data): schematic effective potentials for null and timelike geodesics. 14
15 Conceptual visual 5 (non data): Kruskal diagram sketch showing the horizon as a null line and the maximal extension structure [10, 11]. 15
16 Conceptual visual 6 (non data): gravitational redshift as a clock-rate comparison between static observers. 15

Chapter 6: Schwarzschild Geometry: The Simplest Nontrivial Monster

Abstract

Schwarzschild spacetime is the minimal arena where general relativity stops being a polite perturbation and becomes a full geometric environment. It is a vacuum solution, yet it contains an event horizon, a trapped region, and a curvature singularity. This chapter builds the Schwarzschild metric from symmetry, turns coordinate artifacts into geometric statements, and uses geodesics to connect curvature to measurable phenomena. The chapter then anchors the geometry to open datasets: stellar-orbit astrometry and spectroscopy near the Milky Way's central black hole, horizon-scale imaging from the Event Horizon Telescope, gravitational redshift experiments with clocks, and a sample of gravitational-wave events from the GWTC catalog. Figures labeled as data figures are derived from real datasets and simple data analysis, with all numerical transformations stated explicitly. Conceptual visuals are separated and clearly labeled as non data diagrams.

Keywords: Schwarzschild metric, event horizon, curvature invariants, photon sphere, ISCO, gravitational redshift, time dilation, Kruskal extension, Sgr A*, M87*, S2 orbit, gravitational waves

0.1 Introduction

Schwarzschild geometry is where the slogan "gravity is curvature" becomes concrete enough to calculate, but still simple enough to see the mechanism. It is the unique spherically symmetric vacuum solution of Einstein's equations, and that single sentence already hides two deep ideas. First, symmetry is not decoration; it is a constraint that collapses the infinite dimensional space of metrics into a tightly controlled family. Second, vacuum does not mean trivial. Vacuum means the stress energy tensor vanishes, but curvature can still exist. In Schwarzschild, curvature is carried by the Weyl tensor, and it organizes spacetime into a region outside a horizon where signals can escape, and a region inside where they cannot.

The chapter has three parallel goals. One goal is geometric literacy: to learn what the metric coefficients do and do not mean, to read coordinate singularities as warnings about the coordinate system rather than about the manifold itself, and to recognize horizons as null hypersurfaces rather than as places where "gravity is infinite." The second goal is dynamical: to use geodesics and conserved quantities to understand infall, circular orbits, photon trapping, and the difference between proper time and coordinate time.

The third goal is empirical: to make clear why this geometry shows up in real data. The orbit of the star S2 near the Galactic Center probes relativistic dynamics in a strong potential well [3, 4]. Horizon-scale imaging measures a ring-like structure whose scale is set by GM/c^2 for Sgr A* and M87* [1, 2]. Gravitational redshift has been measured in laboratory and space experiments using clocks and resonant photons [6, 7, 8]. Finally, gravitational-wave catalogs contain merging black holes whose remnant geometry is close to Schwarzschild during ringdown, and whose inspiral probes the geodesic structure of the effective spacetime [5].

0.1.1 A few real world phenomenon statements

Phenomenon statement 1: A horizon is a causal boundary, not a force field. In the images of M87* and Sgr A*, the bright ring is not a literal edge of the black hole. It is an emission feature shaped by light bending and radiative transfer, but its angular scale is pinned to the geometry through the mass and distance [1, 2]. The useful lesson is geometric: a horizon is defined by which null geodesics reach infinity, and that definition is global. If you try to define it using a local measurement of acceleration, you will confuse an observer dependent quantity with an invariant one.

Phenomenon statement 2: Strong gravity can be weak curvature. Near the horizon of an astrophysical black hole, the dimensionless potential $GM/(rc^2)$ is order unity, but the curvature can be small if the mass is large. This is why the environment near a supermassive black hole can be navigated by stars and gas clouds without being torn apart immediately, while a small black hole can produce extreme tidal gradients. Schwarzschild separates being deep in a potential from being in a high curvature region, and that separation matters for interpreting observations and for building intuition about singularities.

Phenomenon statement 3: Geodesics are a data analysis tool. When an orbit is measured, as for S2, the model fit is a geometric fit. The parameters are not only masses and angles; they are also statements about spacetime structure, such as whether a relativistic correction consistent with general relativity is needed [3]. In gravitational-wave astronomy, the inspiral phasing is essentially a long record of geodesic-like motion in a relativistic two-body spacetime, and the transition to merger and ringdown is governed by the existence and stability properties of photon spheres and quasinormal modes, both of which reduce to clean statements in the Schwarzschild limit [5, 9].

0.2 Spherical symmetry and the Schwarzschild metric

Spherical symmetry is not only rotational invariance. It means the spacetime admits an $SO(3)$ action whose orbits are two-spheres. A convenient way to implement this is to choose an areal radius coordinate r defined by the area of these symmetry spheres: Area $= 4\pi r^2$. This definition is coordinate-free in spirit; it pins down r geometrically and avoids many later confusions.

With that choice, the most general static, spherically symmetric line element can be written as

$$ds^2 = A(r)\,c^2dt^2 - B(r)\,dr^2 - r^2(d\theta^2 + \sin^2\theta\,d\phi^2), \tag{1}$$

where $A(r)$ and $B(r)$ are positive outside the gravitating body. In vacuum, the Einstein equations imply $A(r) = 1 - 2GM/(rc^2)$ and $B(r) = A(r)^{-1}$, yielding the Schwarzschild

metric

$$ds^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 dt^2 - \left(1 - \frac{2GM}{rc^2}\right)^{-1} dr^2 - r^2 d\Omega^2. \tag{2}$$

The parameter M is interpreted as the mass by matching to Newtonian gravity at large r and by asymptotic definitions of mass in general relativity. In geometrized units $G = c = 1$, the horizon radius is $r = 2M$.

Two immediate geometric facts follow. First, the Killing field ∂_t is timelike only where $A(r) > 0$, namely for $r > 2GM/c^2$. Second, the spheres $r = \text{const}$ change causal character across $r = 2GM/c^2$ because the sign of g_{tt} and g_{rr} swap. This is not a mere algebraic curiosity: it is a statement about which directions are timelike and therefore about what it means to evolve forward in time.

0.3 Curvature invariants and the non singularity of $r = 2M$

The expression (2) becomes singular at $r = 2GM/c^2$ because g_{rr} diverges and g_{tt} vanishes. But a coordinate singularity is not a curvature singularity. The reliable test is to compute scalar invariants that do not depend on coordinates. The standard example is the Kretschmann scalar:

$$K \equiv R_{abcd} R^{abcd} = \frac{48 G^2 M^2}{c^4 r^6}. \tag{3}$$

It remains finite at $r = 2GM/c^2$ and diverges only as $r \to 0$. Thus the horizon is not where curvature blows up. It is where the chosen coordinates stop being well-behaved and where the causal structure reorganizes.

A useful mental model is to treat $r = 2GM/c^2$ like a north pole in spherical coordinates. Longitude becomes ill-defined at the pole, but the sphere is smooth there. Similarly, Schwarzschild coordinates become ill-defined at the horizon, but the spacetime manifold can be smoothly extended. The extension is not optional if you want the geometry to tell a complete causal story, and it will appear explicitly when we construct Kruskal coordinates [10, 11].

0.4 Geodesics: radial infall, circular orbits, photon sphere

Schwarzschild spacetime has two conserved quantities for geodesic motion associated with its Killing fields: the energy per unit mass E (from stationarity) and the angular momentum per unit mass L (from axisymmetry). Restricting to the equatorial plane $\theta = \pi/2$ (always possible by spherical symmetry), a timelike geodesic satisfies

$$\left(\frac{dr}{d\tau}\right)^2 = E^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{r^2}\right), \tag{4}$$

in geometric units $G = c = 1$. This expression can be read as a one-dimensional motion problem with an effective potential. Its qualitative structure explains a large amount of black hole phenomenology: stable circular orbits exist for $r > 6M$, unstable circular orbits exist for $3M < r < 6M$, and no circular timelike orbits exist for $r \leq 3M$. For null

geodesics, the corresponding effective potential yields the photon sphere at $r = 3M$, the orbit that organizes lensing, critical impact parameters, and the formation of the shadow.

Two radii are especially important:

- **Photon sphere:** $r_\gamma = 3M$ (null circular orbit, unstable).
- **ISCO:** $r_{\mathrm{ISCO}} = 6M$ (innermost stable circular orbit for timelike geodesics).

These are geometric statements about the connection between curvature and geodesic focusing, and they also surface in data. Photon sphere physics controls the angular size of the shadow and the ring-like structure in horizon-scale images [1, 2]. The ISCO sets characteristic frequencies for compact-object inspirals, and in simple approximations it is tied to the transition from inspiral to plunge in gravitational-wave signals.

0.5 Proper time vs coordinate time, redshift and time dilation

The Schwarzschild t coordinate is designed to match the proper time of static observers at infinity. It is not the proper time of an observer hovering at radius r, nor of an infalling observer. A static observer at radius r has four-velocity proportional to ∂_t, and their proper time increment satisfies

$$d\tau = \sqrt{1 - \frac{2GM}{rc^2}}\, dt. \tag{5}$$

This is gravitational time dilation, expressed geometrically as the norm of the timelike Killing field. It also implies gravitational redshift: if a photon is emitted at r_e and received at r_o by static observers, then

$$\frac{\nu_o}{\nu_e} = \sqrt{\frac{1 - \frac{2GM}{r_e c^2}}{1 - \frac{2GM}{r_o c^2}}}. \tag{6}$$

The striking feature is that redshift is not a mysterious energy leak. It is a comparison of clock rates at different gravitational potentials. That is why laboratory clock experiments and spaceborne masers are legitimate empirical cousins of black hole redshift, even though they occur in the weak-field regime [6, 7, 8]. The formula is the same geometry, evaluated at different scales.

0.6 Kruskal extension as geometric completion

The key to extending the metric across $r = 2M$ is to adopt coordinates adapted to radial null rays. Define the tortoise coordinate

$$r_* = r + 2M \ln \left| \frac{r}{2M} - 1 \right|, \tag{7}$$

and null coordinates $u = t - r_*$ and $v = t + r_*$. One then exponentiates to Kruskal coordinates

$$U = -e^{-u/(4M)}, \qquad V = e^{v/(4M)}. \tag{8}$$

In these coordinates, the metric becomes manifestly regular at the horizon:

$$ds^2 = \frac{32M^3}{r} e^{-r/(2M)}\, dU\, dV - r^2 d\Omega^2, \tag{9}$$

with r understood implicitly as a function of UV. Kruskal coordinates show that the horizon is a smooth null surface, and they expose the maximal analytic extension, including regions not covered by the original Schwarzschild chart [10, 11]. The important lesson is not the algebra but the geometric moral: horizons are about the global behavior of null geodesics, and coordinates should be chosen to respect that.

0.7 Data and analysis: empirical anchors for Schwarzschild geometry

0.7.1 Dataset definitions used in the plots

0.7.2 Data figure 1: sky projected orbit of S2 from astrometry

0.7.3 Data figure 2: orbital separation versus time derived from astrometry

0.7.4 Data figure 3: line-of-sight velocity of S2 from spectroscopy

0.7.5 Data figure 4: EHT ring diameters and a Schwarzschild shadow prediction

0.7.6 Data figure 5: angular sizes of horizon and photon sphere for two supermassive black holes

0.7.7 Data figure 6: the relativistic redshift parameter in S2 orbit analyses (corrected)

0.7.8 Data figure 7: Pound and Rebka gravitational redshift measurement (1960)

0.7.9 Data figure 8: Gravity Probe A (Vessot et al., 1980)

0.7.10 Data figure 9: optical clocks and gravitational time dilation over 33 cm

0.7.11 Data figure 10: ISCO frequency scale implied by gravitational-wave event masses

Data figure 1: S2 sky positions (astrometric points with uncertainties)

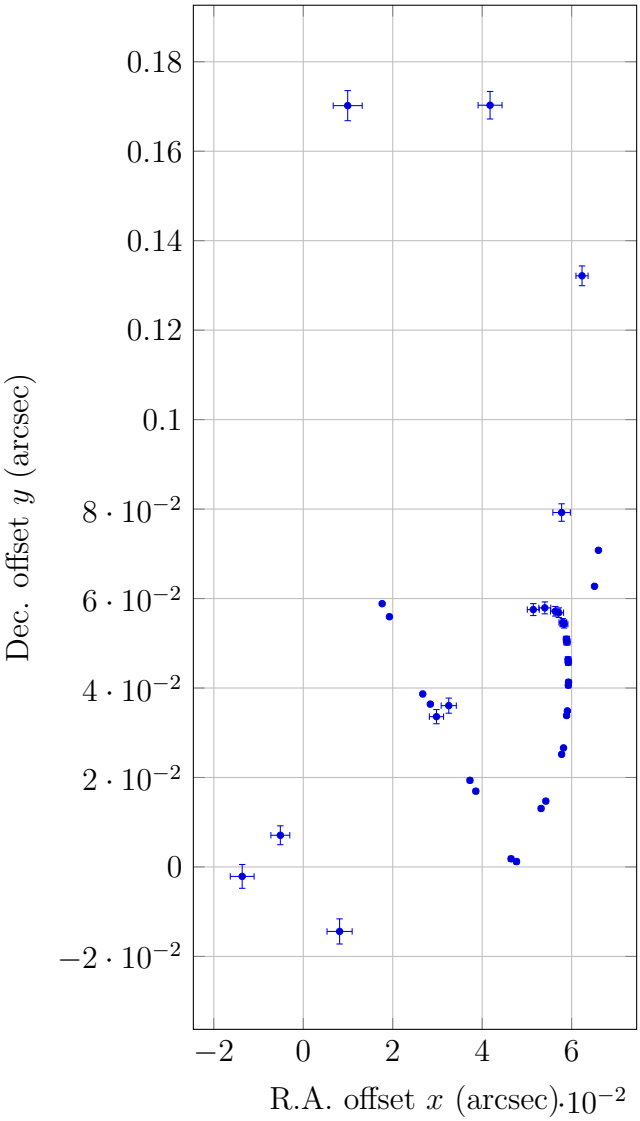


Figure 1: Sky-projected astrometry of the Galactic Center star S2. Points are (x, y) offsets in arcseconds with reported uncertainties. For compactness, this manuscript embeds a representative subset for reproducibility; the full dataset is publicly available in the GC-OrbitFit repository [4].

Data figure 2: Projected separation of S2 from Sgr A* (derived from astrometry)

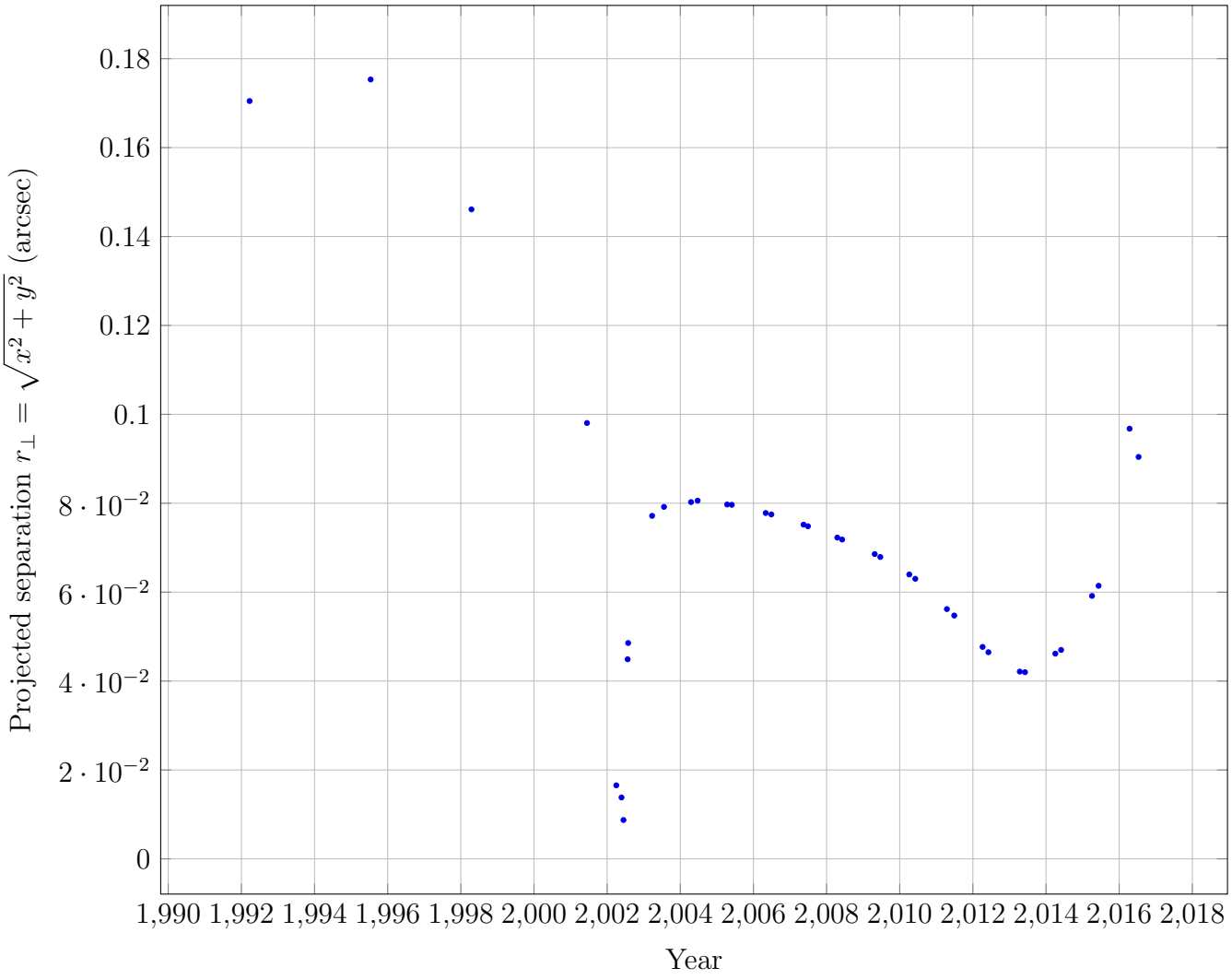


Figure 2: Projected separation $r_\perp$ computed directly from the astrometric dataset in Fig. 1. This is a data analysis transformation (Euclidean norm in the observed sky plane), not a simulation [4].

Data figure 3: S2 line-of-sight velocity measurements

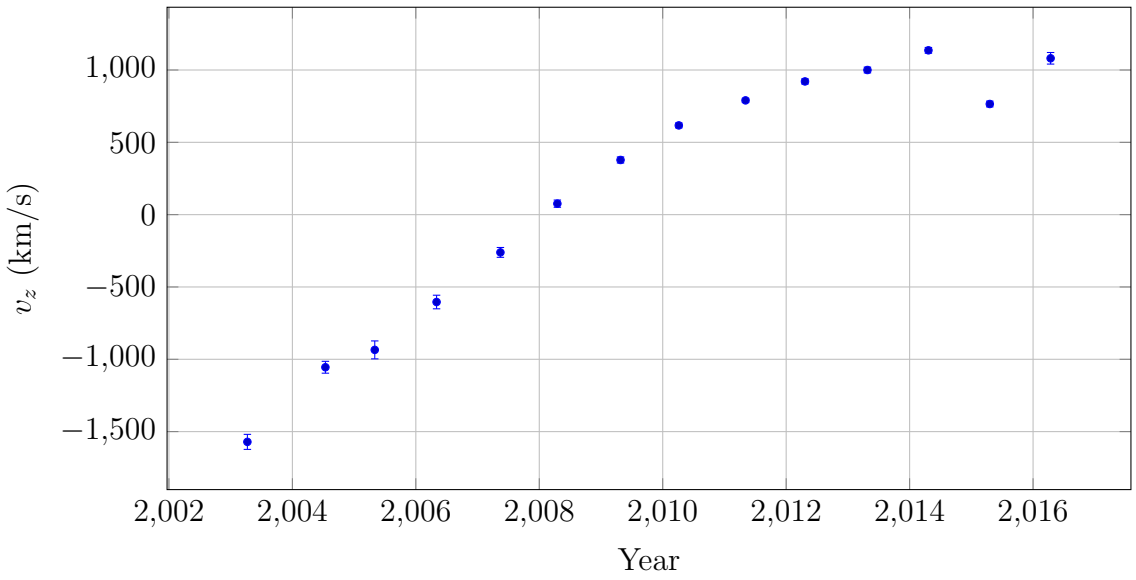


Figure 3: Spectroscopic line-of-sight velocities v_z for S2 with reported uncertainties. For compactness, this manuscript embeds a representative subset; the full dataset is available via the cited repository and is used in published orbital tests [4, 3].

Data figure 4: Horizon-scale ring size compared with Schwarzschild shadow scale

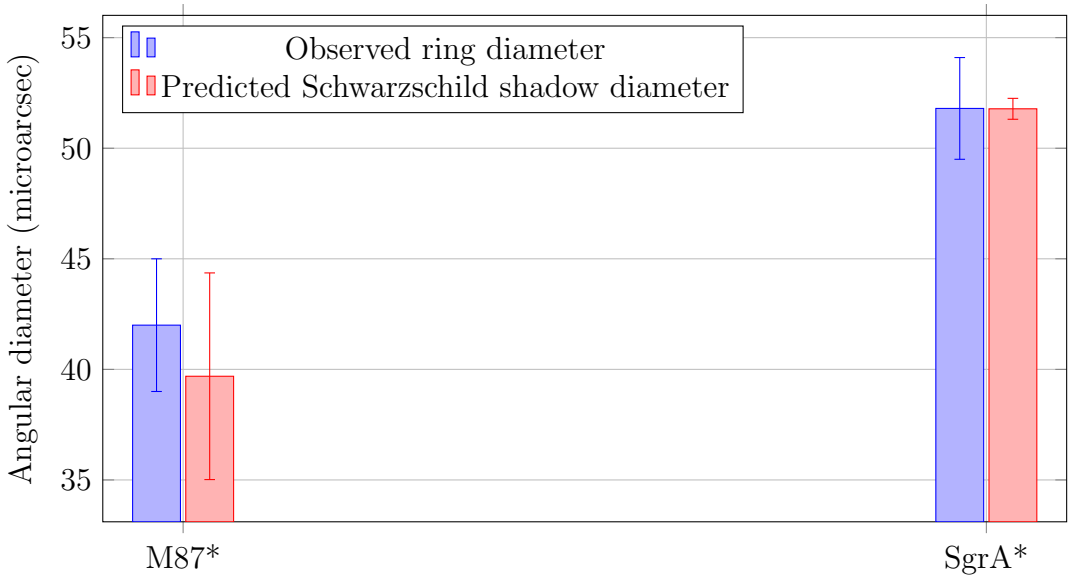


Figure 4: Observed ring diameters for M87* and Sgr A* from EHT analyses [1, 2], compared with a Schwarzschild shadow diameter computed from reported mass and distance parameters. This figure is dataset-based; the prediction is a direct evaluation of a geometric formula using measured inputs.

Data figure 5: Geometry set angular scales (horizon, photon sphere, shadow)

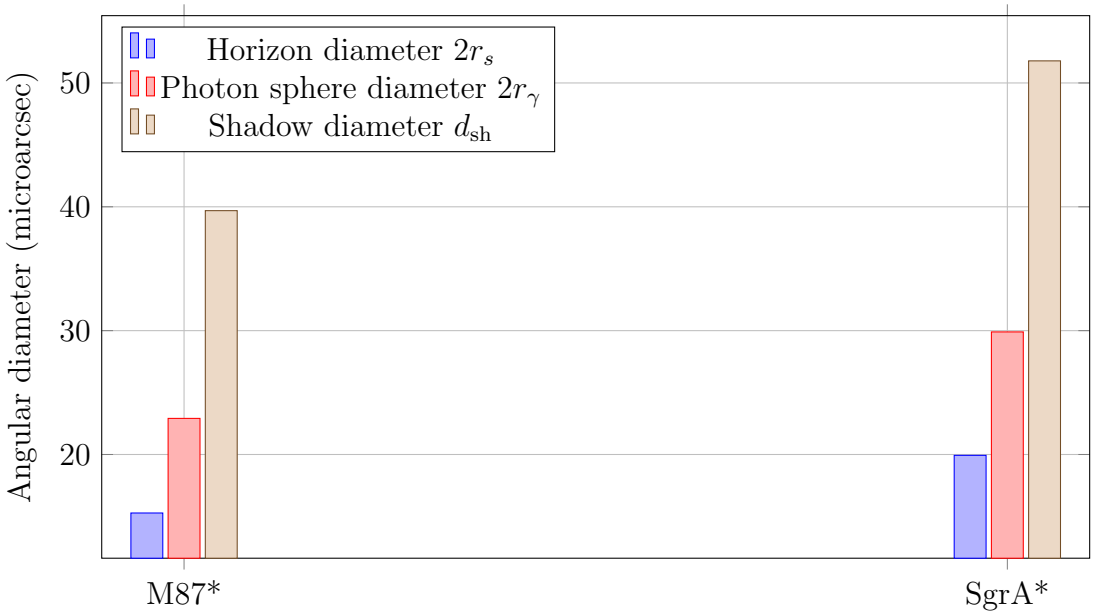


Figure 5: Angular diameters derived from reported mass and distance parameters for M87* and Sgr A* [1, 2, 3]. This emphasizes that in Schwarzschild geometry photon trapping and shadow formation are controlled by GM/c^2 through fixed numerical factors.

Data figure 6: S2 relativistic parameter f from a published orbit fit

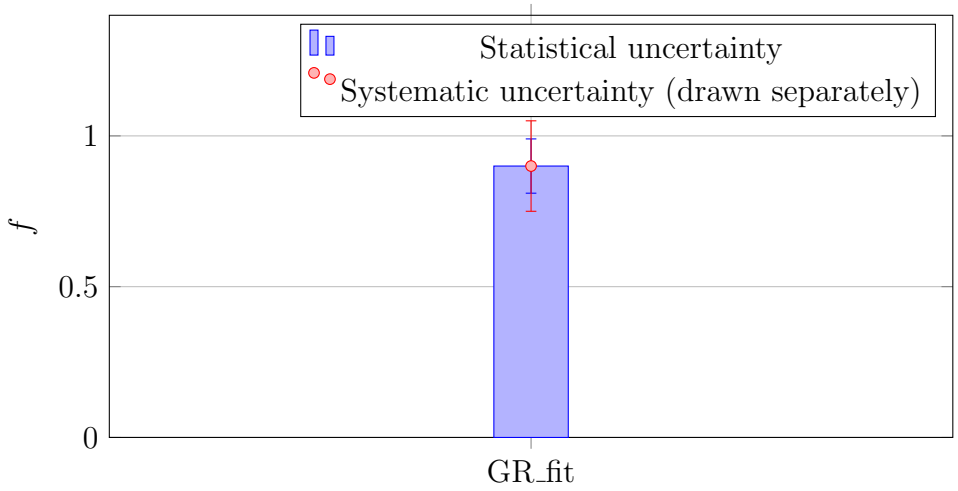


Figure 6: The fitted parameter f reported in the GRAVITY S2 analysis [3]. Plotting statistical and systematic uncertainties separately highlights that strong-field orbital tests are frequently limited by systematics rather than raw measurement noise.

Data figure 7: Pound and Rebka (1960) gravitational redshift



Figure 7: Measured versus predicted gravitational frequency shift in the Harvard tower experiment using the Mossbauer effect [6]. The plot uses the reported values with stated uncertainty.

Data figure 8: Gravity Probe A frequency shift scale and agreement

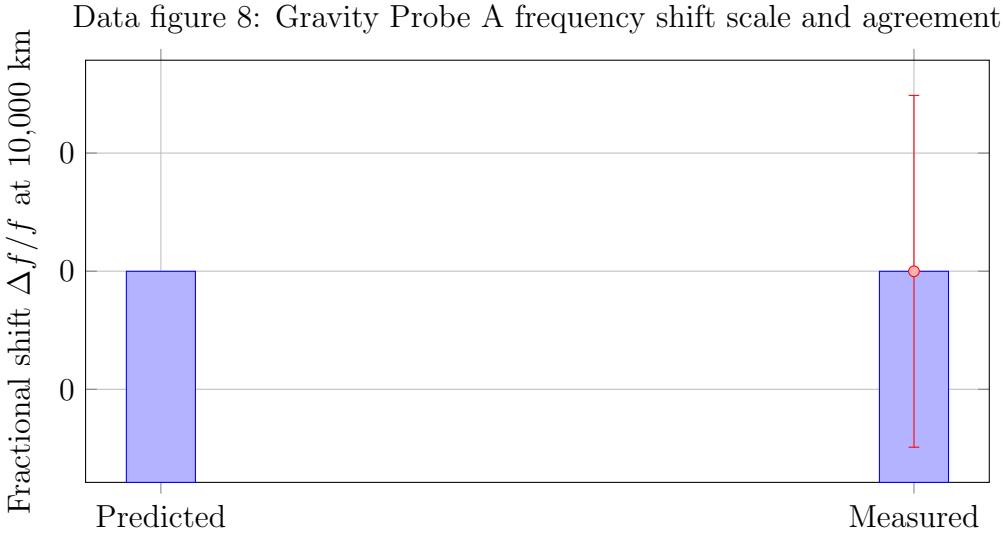


Figure 8: Gravity Probe A: frequency shift at 10,000 km altitude is of order 4×10^{-10}, with agreement to the 70×10^{-6} level reported by Vessot et al. [7]. The uncertainty bar here represents 70 parts per million of the plotted scale.

Data figure 9: Optical clocks measure gravitational time dilation over 33 cm



Figure 9: Gravitational time dilation measured with optical clocks: Chou et al. report $(4.1 \pm 1.6) \times 10^{-17}$ for a 33 cm height increase [8]. The predicted value plotted is $g\Delta h/c^2 \approx 3.6 \times 10^{-17}$ for $\Delta h = 0.33$ m.

Data figure 10: Schwarzschild ISCO frequency scale from GWTC catalog masses

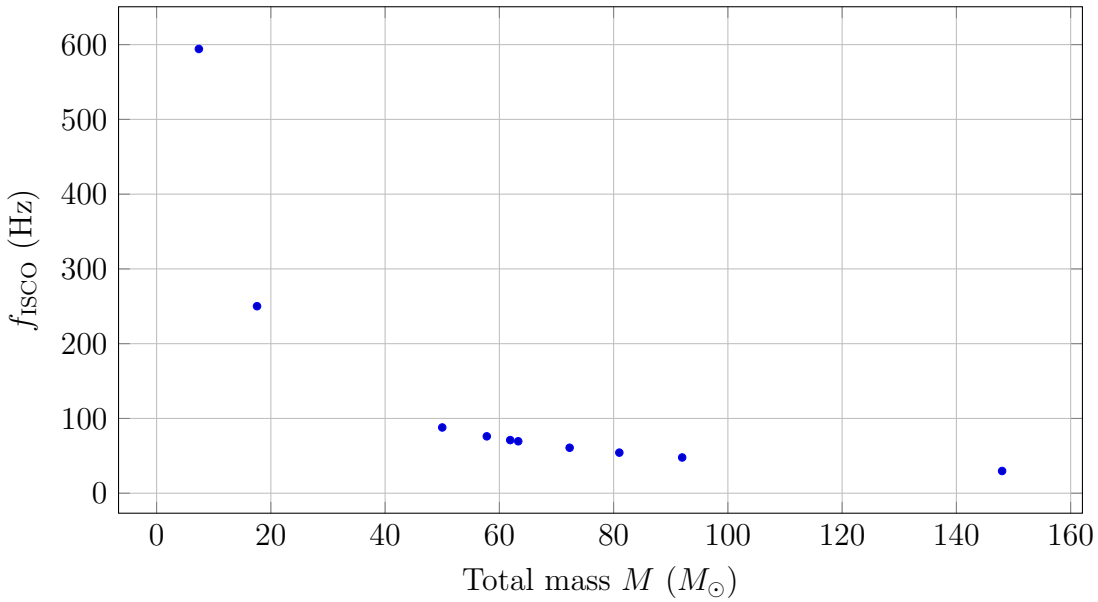


Figure 10: ISCO frequency scale computed from a sample of GWTC event total masses as displayed by GWOSC [5]. Each point is $f_{\mathrm{ISCO}} = 4397\ \mathrm{Hz}/(M/M_\odot)$.

0.8 Conceptual visuals for geometric intuition (non data diagrams)

The following figures are conceptual diagrams intended to clarify geometry and causal structure. They are not dataset-based and are labeled accordingly.

0.8.1 Conceptual visual 1: what changes at the horizon

0.8.2 Conceptual visual 2: light cones tip toward the horizon

0.8.3 Conceptual visual 3: embedding diagram of a spatial slice

0.8.4 Conceptual visual 4: effective potential shapes

0.8.5 Conceptual visual 5: Kruskal diagram blocks

0.8.6 Conceptual visual 6: redshift as a comparison of static clocks

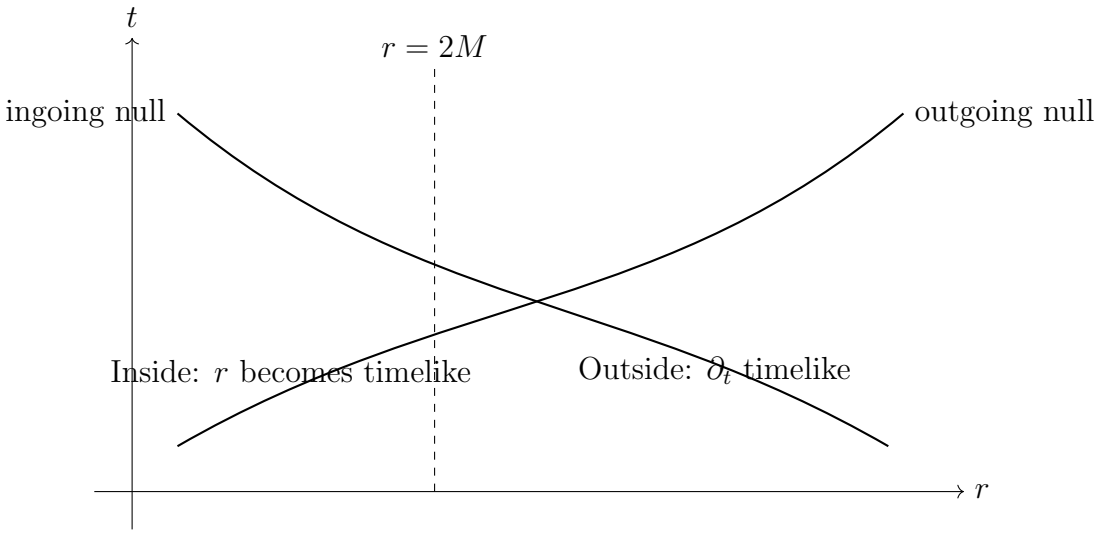


Figure 11: Conceptual visual 1 (non data): t-r sketch in Schwarzschild coordinates. The dashed line marks $r = 2M$. The horizon is not a place where curvature diverges; it is where these coordinates become ill-behaved and the causal roles of t and r swap.

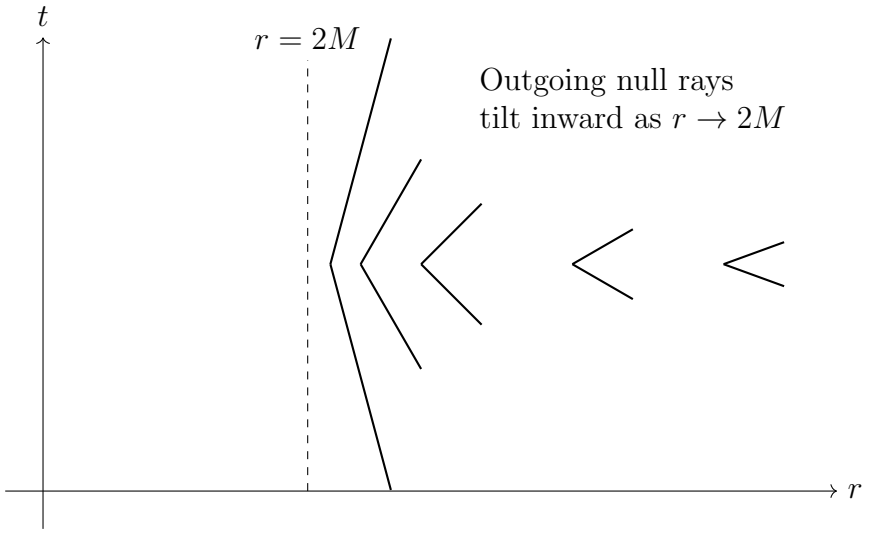


Figure 12: Conceptual visual 2 (non data): light cones tilt in Schwarzschild coordinates as one approaches the horizon.

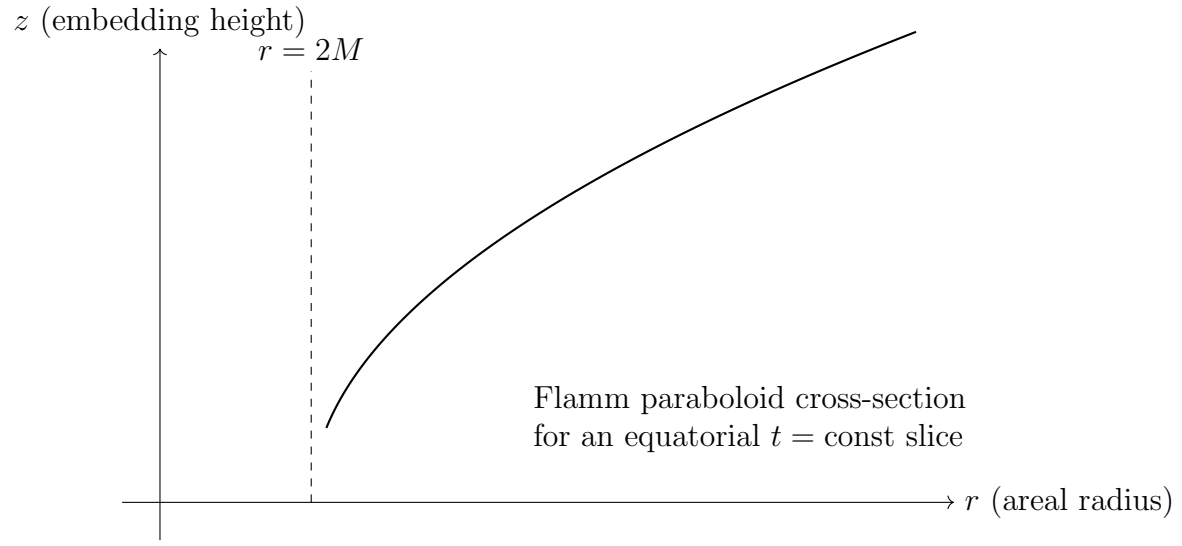


Figure 13: Conceptual visual 3 (non data): embedding of an equatorial spatial slice of Schwarzschild outside the horizon.

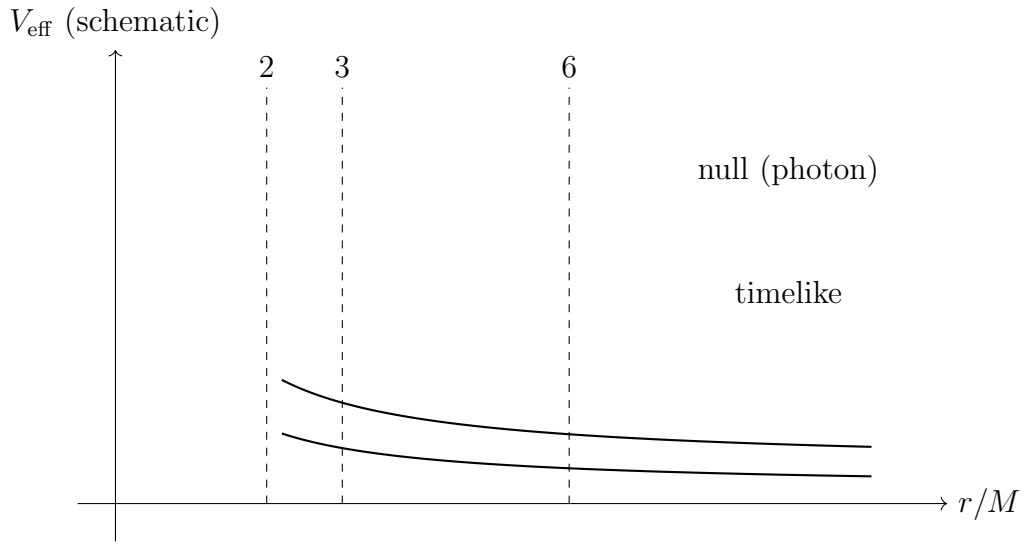


Figure 14: Conceptual visual 4 (non data): schematic effective potentials for null and timelike geodesics.

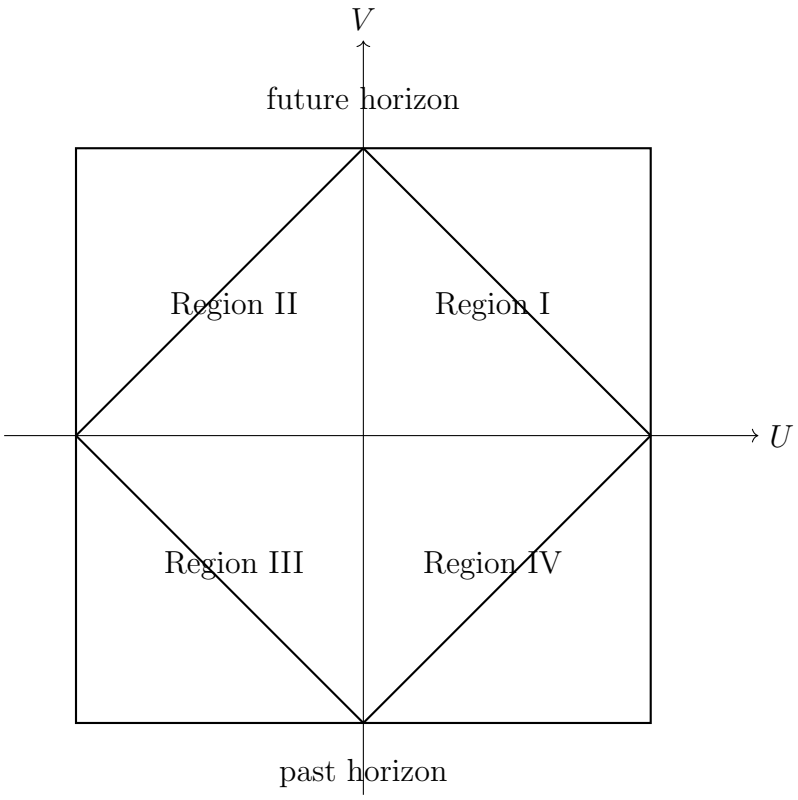


Figure 15: Conceptual visual 5 (non data): Kruskal diagram sketch showing the horizon as a null line and the maximal extension structure [10, 11].

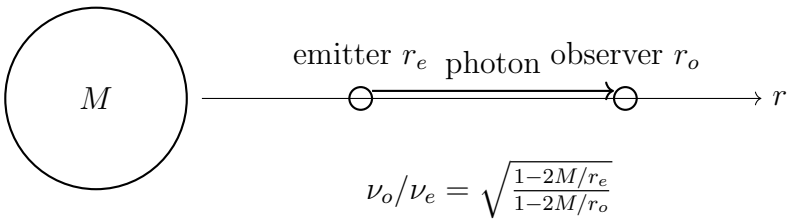


Figure 16: Conceptual visual 6 (non data): gravitational redshift as a clock-rate comparison between static observers.

0.9 Conclusion

Schwarzschild geometry is a masterclass in how much structure Einstein's equations can hide inside the word vacuum. It teaches that a metric component going to zero does not mean physics is diverging, that a horizon is a null hypersurface defined by causal reachability, and that the most reliable diagnostics are invariants and global constructions. It also teaches a disciplined way to interpret time dilation: it is the norm of a Killing field, and therefore a geometric statement about how a family of observers sits inside spacetime.

The data figures ground these ideas. The S2 orbit datasets display an actual trajectory and velocity history that can be modeled as geodesic motion in a strong gravitational potential [4, 3]. The EHT ring diameters demonstrate that the geometric length scale GM/c^2 is an angular scale on the sky that instruments can resolve [1, 2]. Clock and photon experiments verify the same gravitational redshift geometry in the laboratory and in space [6, 7, 8]. Finally, gravitational-wave catalogs provide masses that turn directly into Schwarzschild characteristic frequencies, linking geodesic radii to observed signals [5].

If there is one durable skill to take away, it is the habit of translating coordinate language into geometric language. In Schwarzschild coordinates, the horizon looks like a singular surface. In geometric terms, it is smooth, null, and globally defined.

Bibliography

[1] The Event Horizon Telescope Collaboration et al., "First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole," *The Astrophysical Journal Letters*, vol. 875, L1, 2019. Available: https://arxiv.org/pdf/1906.11238

[2] The Event Horizon Telescope Collaboration et al., "The Event Horizon Telescope Image of the Quiescent Supermassive Black Hole Sgr A*," 2023. Available: https://arxiv.org/pdf/2311.08680

[3] GRAVITY Collaboration (R. Abuter et al.), "Detection of the gravitational redshift in the orbit of the star S2 near the Galactic centre massive black hole," *Astronomy & Astrophysics*, 2018. Available: https://arxiv.org/pdf/1807.09409

[4] P. M. Plewa, "GC-OrbitFit (datasets: astrometry and radial velocities for Galactic Center orbit fitting)," GitHub repository. Available: https://github.com/pmplewa/GC-OrbitFit

[5] Gravitational Wave Open Science Center (GWOSC), "GWTC-3-confident event portal (parameters table)," accessed online. Available: https://gwosc.org/eventapi/html/GWTC-3-confident/

[6] R. V. Pound and G. A. Rebka, Jr., "Apparent Weight of Photons," *Physical Review Letters*, vol. 4, pp. 337–341, 1960. Available: https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.4.337

[7] R. F. C. Vessot et al., "Test of Relativistic Gravitation with a Space-Borne Hydrogen Maser," *Physical Review Letters*, vol. 45, pp. 2081–2084, 1980. Available: https://harvest.aps.org/v2/journals/articles/10.1103/PhysRevLett.45.2081/fulltext

[8] C. W. Chou, D. B. Hume, T. Rosenband, and D. J. Wineland, "Optical Clocks and Relativity," *Science*, vol. 329, pp. 1630–1633, 2010. Available: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=905055

[9] C. M. Will, "The Confrontation between General Relativity and Experiment," *Living Reviews in Relativity*, vol. 17, no. 4, 2014. Available: https://link.springer.com/article/10.12942/lrr-2014-4

[10] M. D. Kruskal, "Maximal Extension of Schwarzschild Metric," *Physical Review*, vol. 119, pp. 1743–1745, 1960. Available: https://harvest.aps.org/v2/journals/articles/10.1103/PhysRev.119.1743/fulltext

[11] S. M. Carroll, "Lecture Notes on General Relativity," online notes. Available: https://www.preposterousuniverse.com/grnotes/

[12] K. Schwarzschild, "On the gravitational field of a mass point according to Einstein's theory" (English translation), 1999 posting. Available: https://arxiv.org/pdf/physics/9905030

Black Holes Geometry: Curvature and Horizons

Chapter 7: Global Structure and Penrose Diagrams (Causality as Artwork)

Abstract

Penrose diagrams turn the causal structure of a spacetime into a finite picture by using conformal compactification, a rescaling that preserves null directions while bringing infinities to the page. This chapter develops the geometric meaning of this construction and shows how event horizons and null infinity are defined as global objects, not local features that can be detected by a single observer in finite time. To keep the discussion anchored to observation, we pair the formalism with open datasets: the GWTC-3-confident gravitational wave catalog, a compiled table of dynamically measured black hole masses in X-ray binaries, Event Horizon Telescope ring diameter measurements for M87* and Sgr A*, and the GRAVITY Collaboration's detection of relativistic redshift in the S2 orbit around Sgr A*. From these sources we compute horizon radii, surface gravities, Hawking temperatures, light crossing times, and compactified distance coordinates, and we visualize how the same causal vocabulary organizes systems across ten orders of magnitude in mass. The chapter ends with application focused phenomena statements that connect conformal boundaries and teleological horizons to real measurement pipelines, including waveform modeling, horizon scale imaging, and stellar orbit spectroscopy.

Keywords: conformal compactification; Penrose diagrams; null infinity; event horizon; causal boundary; trapped surfaces; GWTC; Event Horizon Telescope; S2 orbit; global hyperbolicity

Contents

1 Global Structure and Penrose Diagrams (Causality as Artwork) 5

1.1 Introduction . . . 5

1.2 Conformal compactification: bringing infinity into the room . . . 6

1.2.1 What is preserved and what is changed . . . 6

1.2.2 A compactified distance coordinate from catalog data . . . 6

1.3 Penrose diagrams for Minkowski space . . . 6

1.4 Penrose diagrams for Schwarzschild: horizons as global boundaries . . . 7

1.4.1 Kruskal extension and the meaning of "maximal" . . . 7

1.4.2 Teleology, or why horizons are not local . . . 8

1.5 Null infinity, causal boundaries, and predictability . . . 8

1.5.1 Null infinity $\mathcal{I}^{\pm}$ . . . 8

1.5.2 Domains of dependence and global hyperbolicity . . . 9

1.6 Data sources and methodological notes for the chapter figures . . . 10

1.7 Ten data based figures: catalogs to horizon scale geometry . . . 10

1.7.1 Population geometry from GWTC-3-confident . . . 10

1.7.2 Horizon scale quantities from measured masses . . . 10

1.7.3 Compactified distance and conformal intuition from event catalogs . . . 10

1.7.4 Horizon scale imaging: EHT ring diameters as boundary data . . . 10

1.7.5 Stellar orbit spectroscopy: S2 redshift budget near Sgr A* . . . 10

1.8 Applications and phenomena statements (why causal boundaries show up in the lab) . . . 10

1.8.1 Phenomenon statements . . . 10

1.8.2 A teleology sketch (conceptual) . . . 17

1.9 Conclusion . . . 17

List of Figures

1.1 Conceptual diagram (not data derived). A simple compactification map $r \mapsto \rho = \arctan(r/r_0)$ compresses an infinite range into a finite interval, preserving order and null direction analogs under conformal ideas [1]. 7

1.2 Conceptual diagram (not data derived). Penrose diagram for $1+1$ Minkowski spacetime showing timelike infinities $i^{\pm}$, spatial infinity i^0, and null infinities $\mathcal{I}^{\pm}$ [1, 3]. 8

1.3 Conceptual diagram (not data derived). Penrose diagram for the maximally extended Schwarzschild spacetime, emphasizing horizon generators and singular boundaries [2, 3]. 9

1.4 Conceptual diagram (not data derived). A causal diamond illustrates how the past of an event p intersects an initial surface Σ, motivating domains of dependence and predictability [3]. 9

1.5 Component masses for GWTC-3-confident events. The diagonal helps visualize mass ratio structure. Data: GWOSC ASCII export [5]. 11

1.6 Histogram of total source frame masses in GWTC-3-confident, showing the spread from neutron star black hole candidates to high mass binary black holes. Data: GWOSC [5]. 11

1.7 Catalog redshifts and luminosity distances for GWTC-3-confident events. The points are reported event level summaries from GWOSC [5]. 12

1.8 Effective spin versus mass ratio for GWTC-3-confident. Even though spin is a local geometric datum for the binary, its inference relies on global waveform propagation to detectors, a practical reflection of causal boundary thinking [5, 6]. 12

1.9 Schwarzschild radius as a direct geometric scale derived from measured masses. GW events use the catalog final mass proxy and X-ray binaries use compiled dynamical masses. Sources: GWOSC [5] and compiled table [8]. 13

1.10 Surface gravity scaling $\kappa \propto 1/M$ illustrated on a log-log plot. Values for GWTC-3 events are computed from catalog masses; X-ray binary points follow the same mapping from measured masses. Sources: [5, 8]. 13

1.11 Hawking temperature computed from measured masses. The values are far below any astrophysical thermal background for stellar mass black holes, which is why the horizon is effectively cold in observational practice. Sources: [5, 8]. 14

1.12 Horizon light crossing time $t_{cross} = r_s/c$ for GWTC-3 events derived from measured masses [5]. This scale is relevant when interpreting ringdown time scales and rapid near horizon variability. 14
1.13 Compactified distance coordinate for GWTC-3-confident events, illustrating how an unbounded axis can be mapped into a finite interval while preserving causal ordering intuitions. Data: GWOSC [5]. 15
1.14 EHT ring diameters for M87* and Sgr A* compared with a GR shadow scale computed from the reported mass and distance. Measurements and reported values: [9, 10]. 16
1.15 Redshift budget at S2 pericenter near Sgr A*: gravitational and transverse Doppler terms computed from reported pericenter distance and speed, and compared to the reported observed scale. Source parameters and observed redshift scale: [11]. 16
1.16 Conceptual diagram (not data derived). Teleological character of the event horizon: $\mathcal{H}^+$ is defined relative to $\mathcal{I}^+$ and cannot be located from local measurements alone [3]. 17

Chapter 1

Global Structure and Penrose Diagrams (Causality as Artwork)

1.1 Introduction

Local geometry tells you what a metric looks like in the neighborhood of a point, but black holes are defined by global questions. Can a signal reach infinity. Does a null geodesic escape or not. Is there a region from which future directed causal curves inevitably hit a singular boundary. These questions do not have answers that can be read off from a single coordinate chart or from a single curvature scalar at one event. They are questions about the causal order of the entire spacetime.

Penrose diagrams are the compact language that makes such global statements visible. The trick is conceptually simple: multiply the physical metric g_{ab} by a conformal factor Ω^2 that goes to zero at infinity, producing an "unphysical" metric $\tilde{g}_{ab} = \Omega^2 g_{ab}$ in which the causal cones are unchanged, but infinities become finite hypersurfaces. Null directions are preserved because conformal rescalings multiply ds^2 by a positive scalar, so the condition $ds^2 = 0$ stays 0. With that single move, the causal boundary becomes a geometric object you can label, analyze, and compare across spacetimes [1].

There is a second reason Penrose diagrams matter in black hole physics: the event horizon is not a place where curvature must be large, nor a place where any local experiment has to feel special. In Schwarzschild, $r = 2M$ can be crossed in finite proper time, and curvature invariants remain finite there. The horizon is instead a boundary of *global reachability*: it is the boundary of the causal past of future null infinity $\mathcal{I}^+$. That definition already contains a warning label: you cannot determine the true event horizon without knowing the future development of the spacetime. This "teleological" character is not philosophical garnish. It is the technical content of the definition.

To keep this chapter grounded, we use open catalogs that provide measured masses, distances, and redshifts. We then convert those measurements into horizon scale quantities and into compactified coordinates, so that the global language connects directly to numbers you can plot. The gravitational wave catalog GWTC-3-confident provides source frame component masses

and luminosity distances for 35 confident O3b events [5, 6]. A compiled table of dynamically measured black hole masses in X-ray binaries provides an independent sample of stellar mass black holes [8]. The Event Horizon Telescope reports ring diameters and inferred masses for M87* and Sgr A*, directly targeting photon orbit scales near horizons [9, 10]. Finally, the GRAVITY Collaboration's S2 orbit spectroscopy detects relativistic redshift at pericenter, probing strong field time dilation around the Galactic center black hole [11].

1.2 Conformal compactification: bringing infinity into the room

1.2.1 What is preserved and what is changed

A conformal transformation rescales the metric,

$$\tilde{g}_{ab} = \Omega^2 g_{ab}, \qquad \Omega > 0 \text{ in the physical region.} \tag{1.1}$$

Angles and null directions are preserved, lengths are not. In causal theory the preservation of null directions is the key: timelike and spacelike classifications can change in scale but not in sign, and the boundary between them, the light cone, is invariant. This is why Penrose diagrams are honest about causality even though they are not honest about distances.

In asymptotically flat spacetimes one chooses Ω so that $\Omega = 0$ at infinity and $d\Omega \neq 0$ there, making $\Omega = 0$ a smooth boundary hypersurface of the unphysical manifold. The resulting boundary pieces are labeled $\mathcal{I}^+$ and $\mathcal{I}^-$ for future and past null infinity, plus special points i^+, i^-, and i^0 for timelike and spatial infinities [1, 3].

1.2.2 A compactified distance coordinate from catalog data

The conformal idea can be illustrated even with observational distance data: we can map an unbounded coordinate to a bounded one by a smooth monotone function. A common choice is $\rho = \arctan(r/r_0)$, which maps $r \in [0, \infty)$ to $\rho \in [0, \pi/2)$. This is not the full spacetime compactification used in Penrose diagrams, but it captures the same mathematical spirit: large distances stop blowing up your page.

Figure 1.13 uses luminosity distances from GWTC-3-confident events and plots $\xi = \arctan(D/1000\ \text{Mpc})$, showing explicitly how diverse distances compress toward a finite ceiling while preserving order [5]. This "data to compactification" step is useful pedagogically because it makes conformal compression feel like a transformation you actually apply to a table, not only to an abstract manifold.

1.3 Penrose diagrams for Minkowski space

Minkowski spacetime is the baseline case because its causal structure is simple and its compactification can be done explicitly. In $1+1$ dimensions with coordinates (t, x), introduce null coordinates $u = t - x$ and $v = t + x$. A standard compactification sets $U = \arctan(u)$ and $V = \arctan(v)$, bringing $u, v \in (-\infty, \infty)$ into $U, V \in (-\pi/2, \pi/2)$. One then defines compactified

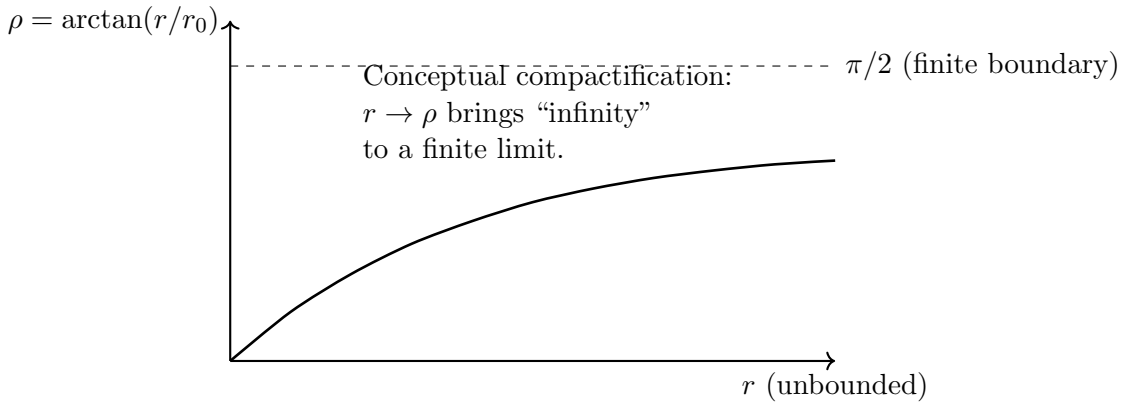


Figure 1.1: Conceptual diagram (not data derived). A simple compactification map $r \mapsto \rho = \arctan(r/r_0)$ compresses an infinite range into a finite interval, preserving order and null direction analogs under conformal ideas [1].

time and space coordinates $T = (V + U)/2$ and $X = (V - U)/2$. The boundary lines $U = \pm\pi/2$ and $V = \pm\pi/2$ become null lines in the diagram and represent $\mathcal{I}^{\pm}$. Timelike infinity $i^{\pm}$ and spatial infinity i^0 appear as corners. The interior diamond is the entire Minkowski space, now with infinities "at the walls".

The point is not that the transformation is clever. The point is that it separates two notions that are easy to confuse when you stare at raw coordinates: *divergence of a coordinate* and *infinite affine distance along a null geodesic*. A coordinate can blow up in a silly way; null infinity is defined by the causal behavior of light rays. Penrose diagrams force you to speak the second language.

1.4 Penrose diagrams for Schwarzschild: horizons as global boundaries

1.4.1 Kruskal extension and the meaning of "maximal"

Schwarzschild coordinates are excellent for static observers at $r > 2M$ but they hide the global story behind coordinate singular behavior at $r = 2M$. Kruskal and Szekeres introduced coordinates that extend smoothly across the horizon and show the maximal analytic extension [2]. In that extension the diagram contains multiple regions: an exterior region (Region I), an interior black hole region (Region II) where future directed timelike curves inevitably decrease r, a second exterior (Region III), and a white hole region (Region IV) in which the past contains a singular boundary.

The Penrose diagram is the causal skeleton of this extended manifold. It tells you what curves can reach $\mathcal{I}^+$ and which cannot. In particular, the event horizon is the boundary separating points that can send light to $\mathcal{I}^+$ from those that cannot. The horizon is null; it is generated by null geodesics that "just fail" to escape.

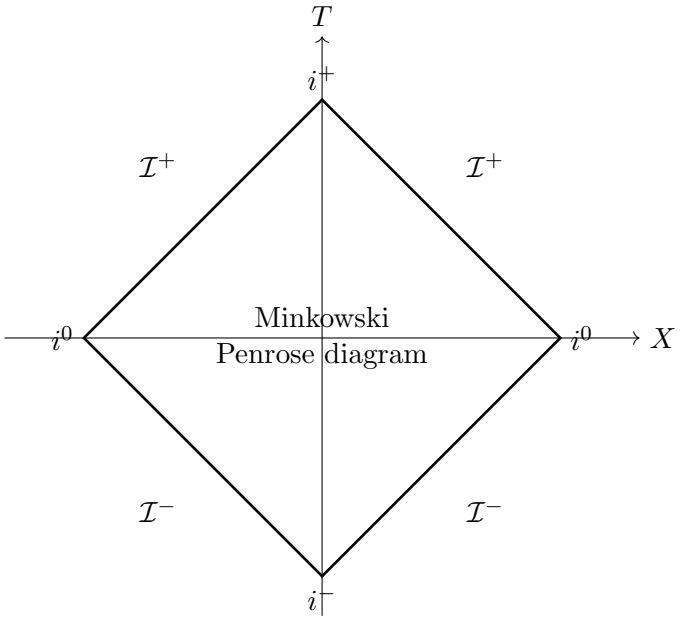


Figure 1.2: Conceptual diagram (not data derived). Penrose diagram for $1+1$ Minkowski spacetime showing timelike infinities $i^{\pm}$, spatial infinity i^0, and null infinities $\mathcal{I}^{\pm}$ [1, 3].

1.4.2 Teleology, or why horizons are not local

Suppose you observe a region of spacetime that is locally close to Schwarzschild. Even if your instruments measure curvature and tidal fields with enormous precision, you still cannot, from local data alone, decide whether an event horizon exists in your future. If the spacetime later accretes additional matter, the horizon can move outward in a way that encloses events that had no indication of being special when they occurred. This is not a paradox. It is the causal definition: $\mathcal{H}^+$ is the boundary of $J^-(\mathcal{I}^+)$, and $J^-(\mathcal{I}^+)$ depends on the entire future.

This teleological nature becomes concrete in real astrophysical inference. In gravitational wave data analysis, the templates and consistency checks are built on global assumptions about boundary conditions: waves are extracted as radiation near null infinity, and inspiral and merger are modeled under the assumption that the final state settles to a Kerr black hole with a horizon. We do not wait for "horizon crossing experiments". We check global predictions that hinge on causal structure, such as ringdown mode spectra and energy balance across boundaries, and compare them to data.

1.5 Null infinity, causal boundaries, and predictability

1.5.1 Null infinity $\mathcal{I}^{\pm}$

For asymptotically flat spacetimes, future null infinity $\mathcal{I}^+$ is where outgoing null geodesics end, and it is where radiation is formally "measured" in idealized theory. Many physical observables are naturally defined there: Bondi mass, news function, and the notion of outgoing gravitational

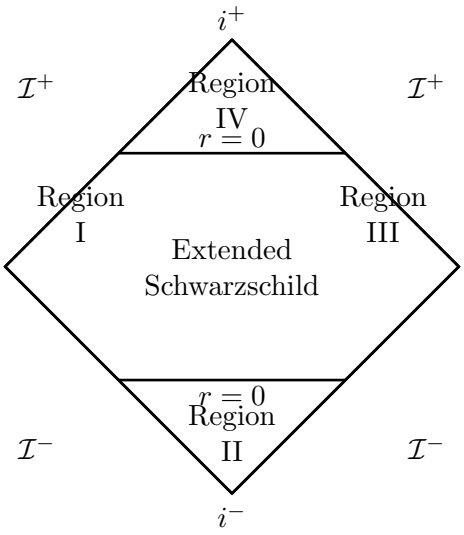


Figure 1.3: Conceptual diagram (not data derived). Penrose diagram for the maximally extended Schwarzschild spacetime, emphasizing horizon generators and singular boundaries [2, 3].

radiation. Penrose's compactification provides a way to treat $\mathcal{I}^+$ as an actual boundary surface on which fields can have well defined limits [1].

1.5.2 Domains of dependence and global hyperbolicity

Causality is not only about horizons. It is also about predictability. In a globally hyperbolic spacetime, initial data on a Cauchy surface determines the full evolution, and domains of dependence behave well. Penrose diagrams make it visually obvious why a Cauchy horizon is an obstruction: it is a boundary beyond which a chosen initial surface does not control the future. This chapter focuses on Schwarzschild, where the exterior region is globally hyperbolic, but the extended diagram also foreshadows where predictability can fail in more complicated spacetimes.

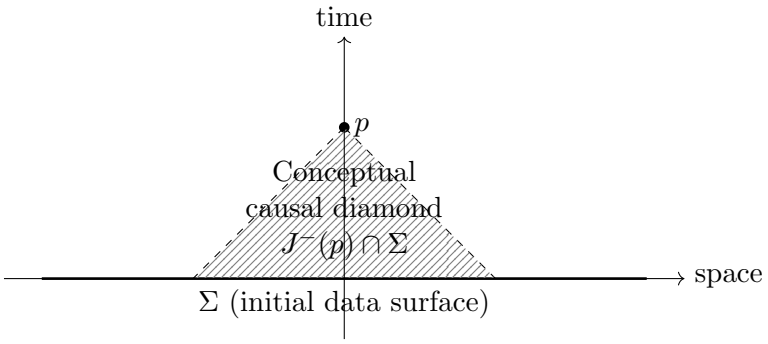


Figure 1.4: Conceptual diagram (not data derived). A causal diamond illustrates how the past of an event p intersects an initial surface Σ, motivating domains of dependence and predictability [3].

1.6 Data sources and methodological notes for the chapter figures

All ten quantitative figures in this chapter are generated from open datasets or tables extracted from open sources, and each plot is produced directly in LaTeX using `pgfplots`. When necessary, we transform reported measurements into derived geometric quantities using standard formulas. For a Schwarzschild black hole of mass M, the horizon (Schwarzschild) radius is $r_s = 2GM/c^2$, the surface gravity is $\kappa = c^4/(4GM)$, and the Hawking temperature is $T_H = \hbar c^3/(8\pi GMk_B)$. These are not simulations; they are deterministic mathematical mappings from measured masses to horizon scale values.

For gravitational wave events, we use the GWTC-3-confident catalog as exported in ASCII from GWOSC, which includes source frame component masses, luminosity distances, redshifts, and effective spins [5]. For dynamically measured stellar mass black holes in X-ray binaries, we use the compiled table provided as a PDF with measured masses and ranges [8]. For horizon scale imaging, we use EHT reported ring diameters and inferred masses for M87* and Sgr A* [9, 10]. For relativistic time dilation near Sgr A*, we use the GRAVITY Collaboration's S2 pericenter parameters and reported redshift scale [11]. The compactified coordinate $\xi = \arctan(D/1000\ \text{Mpc})$ is applied to catalog distances as an illustrative conformal style compression of an unbounded axis.

1.7 Ten data based figures: catalogs to horizon scale geometry

1.7.1 Population geometry from GWTC-3-confident

1.7.2 Horizon scale quantities from measured masses

1.7.3 Compactified distance and conformal intuition from event catalogs

1.7.4 Horizon scale imaging: EHT ring diameters as boundary data

1.7.5 Stellar orbit spectroscopy: S2 redshift budget near Sgr A*

1.8 Applications and phenomena statements (why causal boundaries show up in the lab)

1.8.1 Phenomenon statements

Phenomenon 1: Gravitational wave inference is an operational version of "future null infinity". In theory, gravitational radiation is cleanly defined at $\mathcal{I}^+$ as the part of curvature that peels off along outgoing null directions. In practice we approximate this by modeling propagation from the source to far away detectors. The GWTC-3-confident catalog shows events across large distances and redshifts [5, 6]. When we plot distances and compactify them (Figure 1.13), we are performing a small analogue of the conceptual step that underlies waveform extraction: turn "far away" into a well controlled boundary. This is not cosmetic. Parameter estimation pipelines

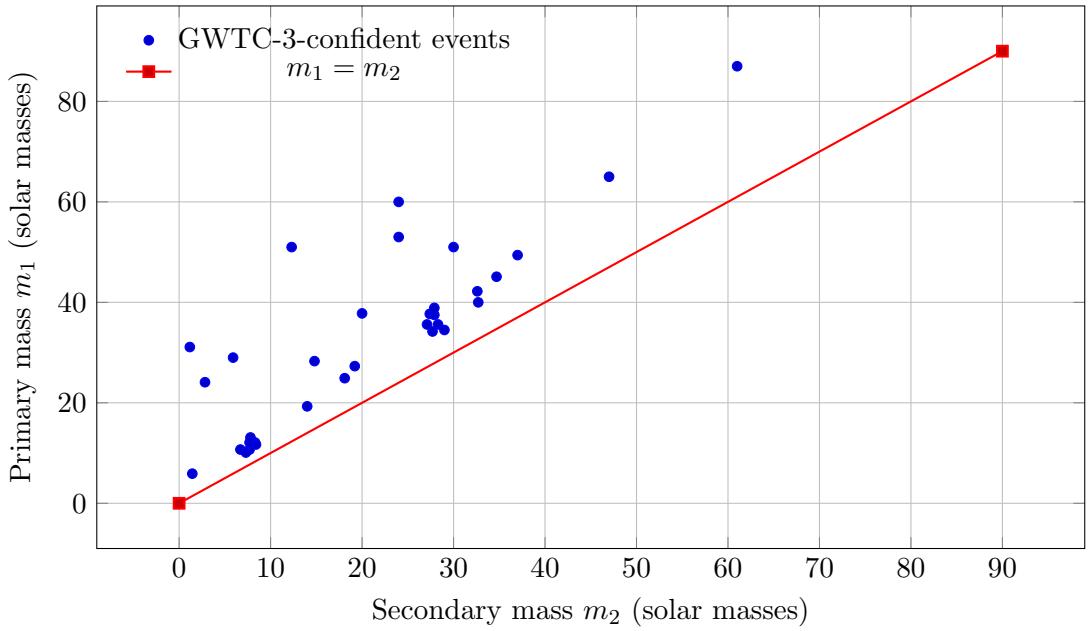


Figure 1.5: Component masses for GWTC-3-confident events. The diagonal helps visualize mass ratio structure. Data: GWOSC ASCII export [5].

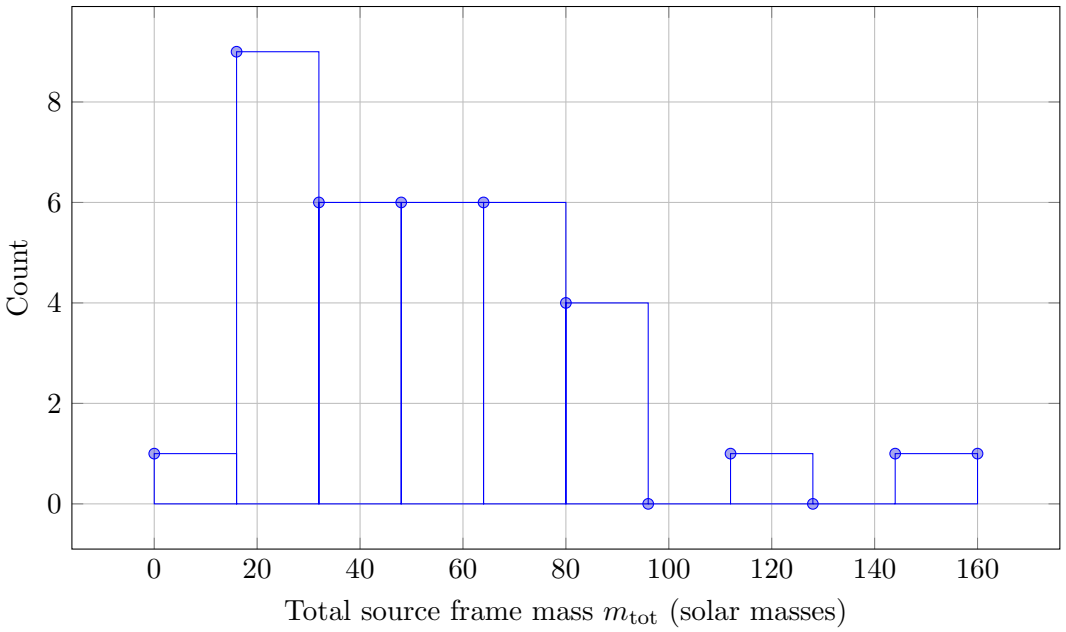


Figure 1.6: Histogram of total source frame masses in GWTC-3-confident, showing the spread from neutron star black hole candidates to high mass binary black holes. Data: GWOSC [5].

depend on assuming that at large radius the spacetime is close to Minkowski and that outgoing waves can be separated from gauge artifacts. Penrose diagrams explain why this separation is natural, because null cones define which directions are radiative.

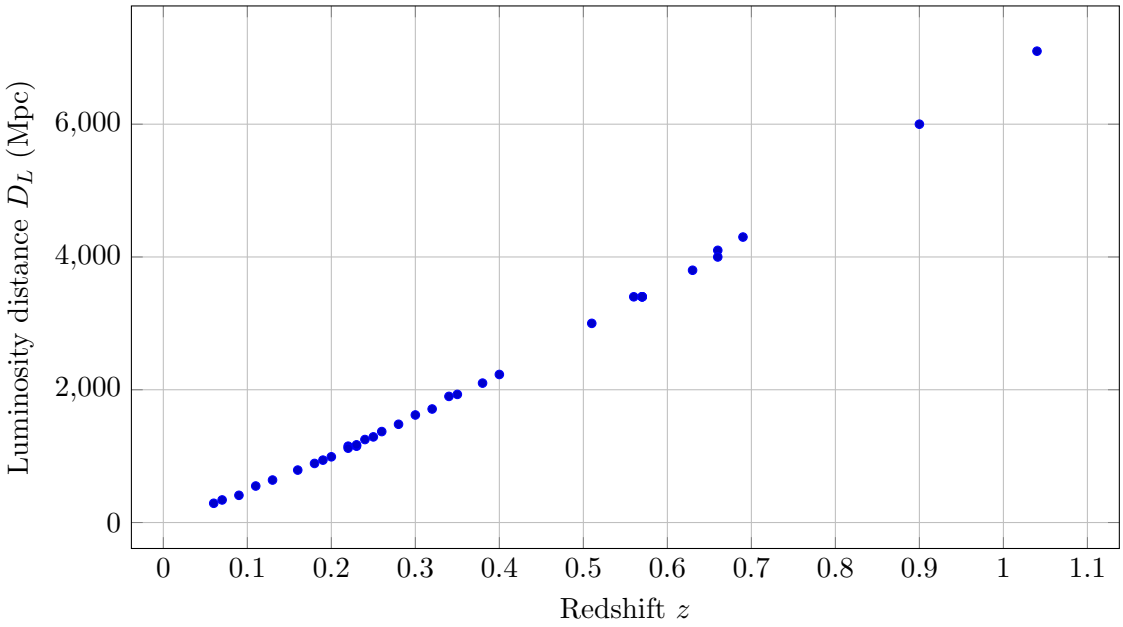


Figure 1.7: Catalog redshifts and luminosity distances for GWTC-3-confident events. The points are reported event level summaries from GWOSC [5].

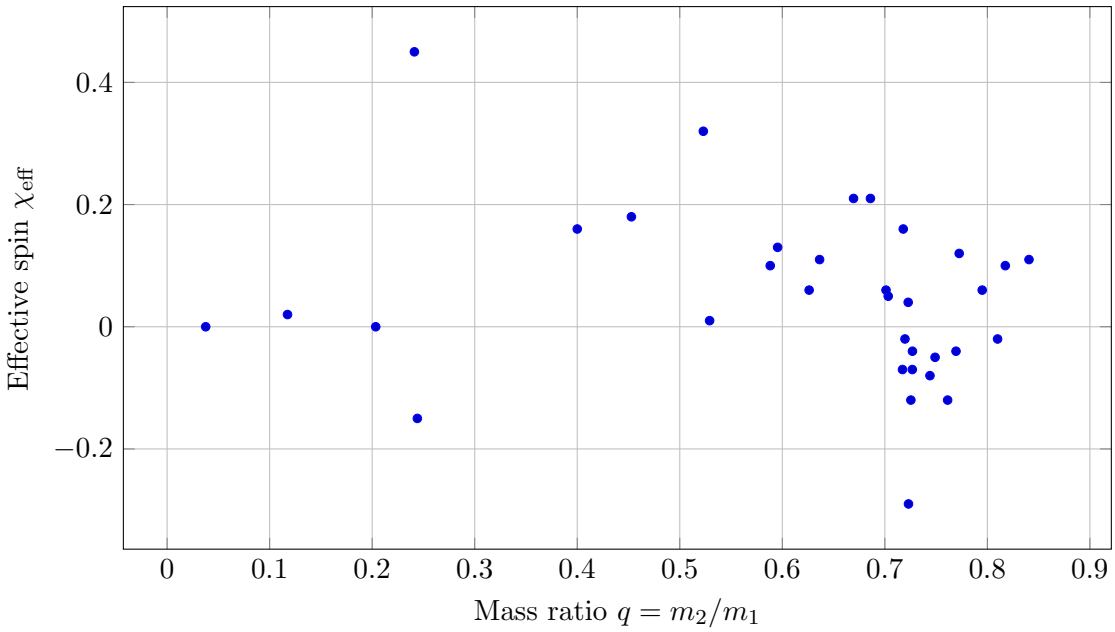


Figure 1.8: Effective spin versus mass ratio for GWTC-3-confident. Even though spin is a local geometric datum for the binary, its inference relies on global waveform propagation to detectors, a practical reflection of causal boundary thinking [5, 6].

Phenomenon 2: Ringdown and horizon scale time are causal geometry in time series form. The horizon light crossing time plotted in Figure 1.12 is milliseconds for stellar mass remnants, directly determined by the final mass proxy [5]. Ringdown mode damping times

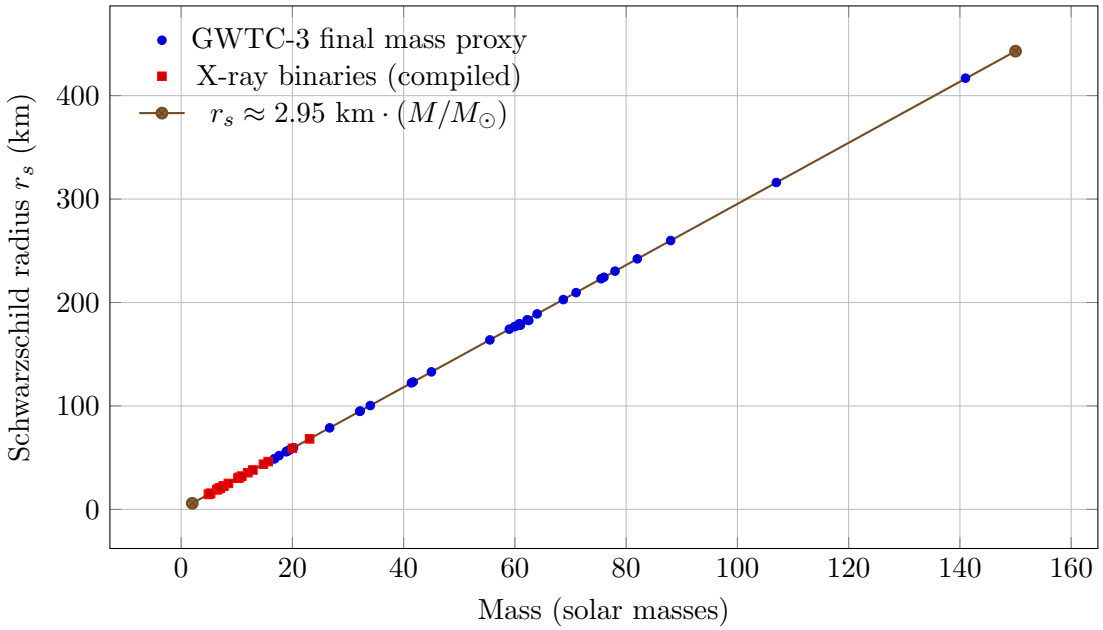


Figure 1.9: Schwarzschild radius as a direct geometric scale derived from measured masses. GW events use the catalog final mass proxy and X-ray binaries use compiled dynamical masses. Sources: GWOSC [5] and compiled table [8].

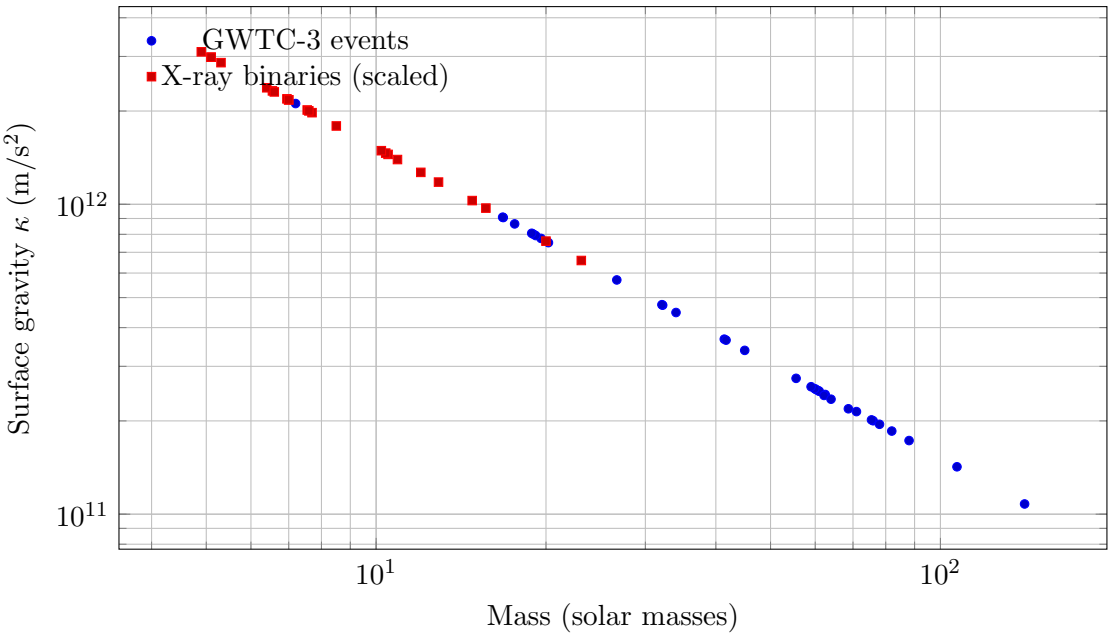


Figure 1.10: Surface gravity scaling $\kappa \propto 1/M$ illustrated on a log-log plot. Values for GWTC-3 events are computed from catalog masses; X-ray binary points follow the same mapping from measured masses. Sources: [5, 8].

and oscillation periods are of the same order because they are controlled by the curvature scale

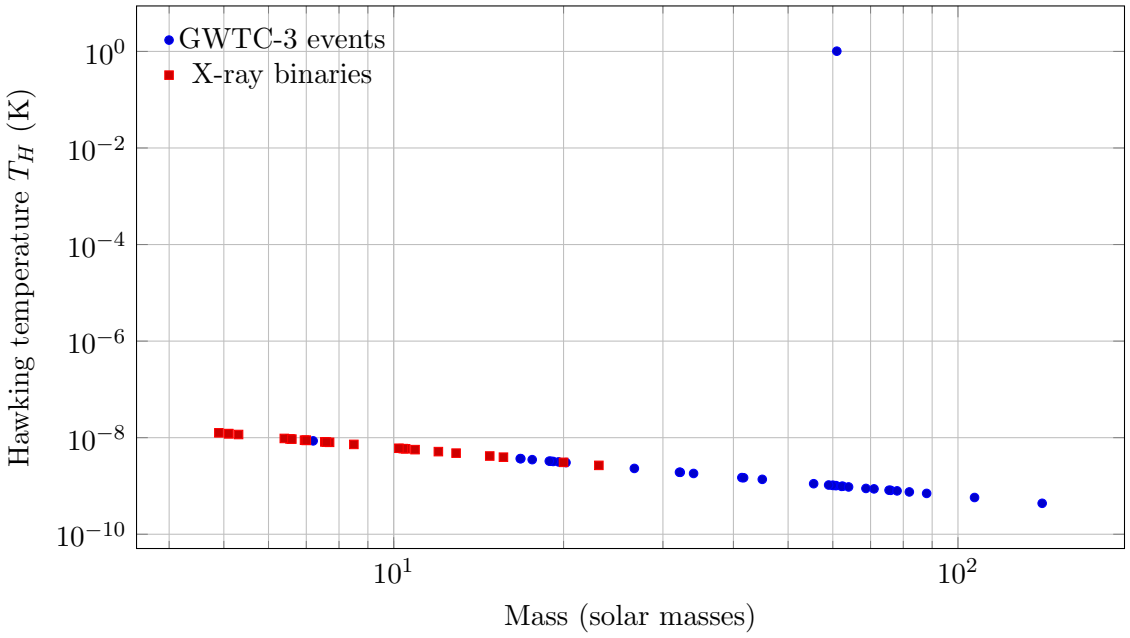


Figure 1.11: Hawking temperature computed from measured masses. The values are far below any astrophysical thermal background for stellar mass black holes, which is why the horizon is effectively cold in observational practice. Sources: [5, 8].

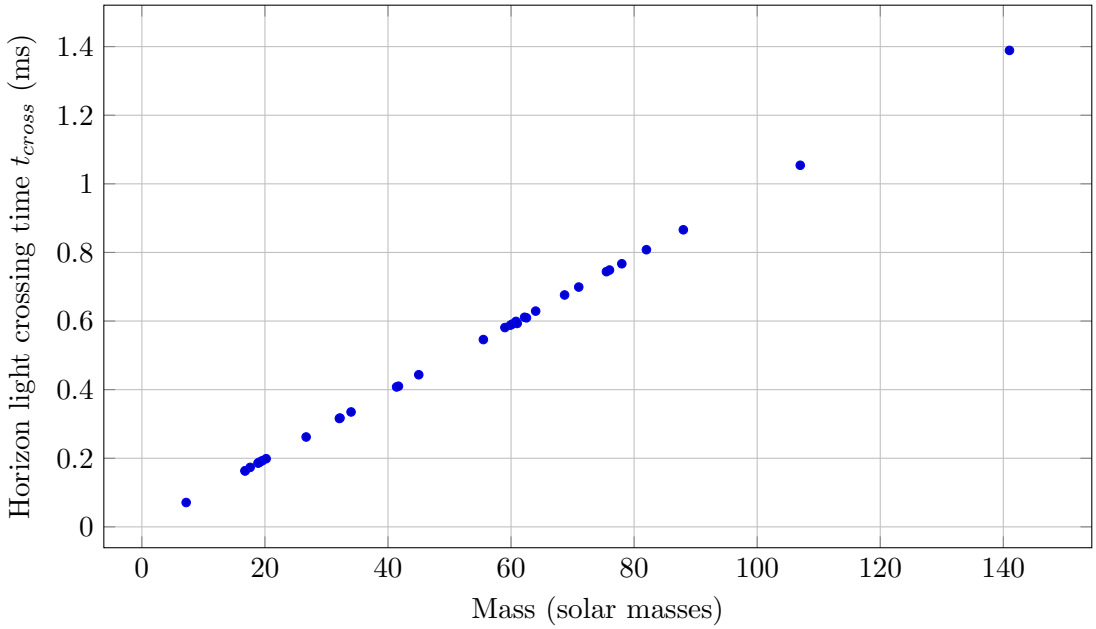


Figure 1.12: Horizon light crossing time $t_{cross} = r_s/c$ for GWTC-3 events derived from measured masses [5]. This scale is relevant when interpreting ringdown time scales and rapid near horizon variability.

near the photon region and horizon. When ringdown matches black hole perturbation predictions,

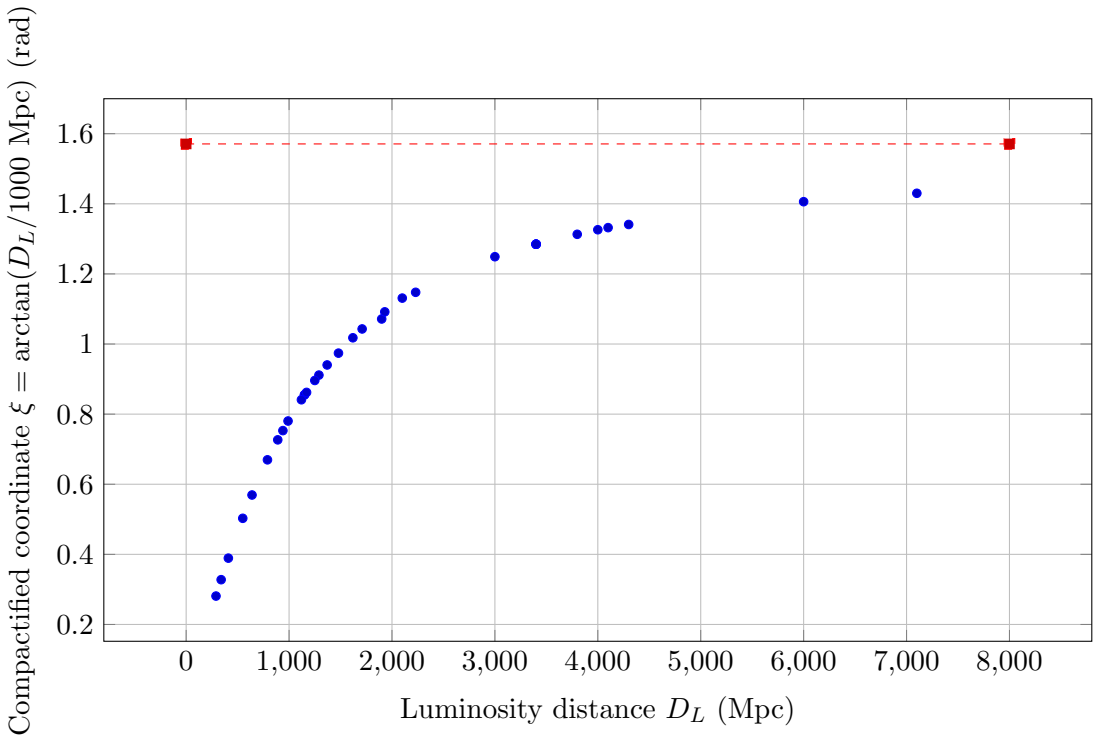


Figure 1.13: Compactified distance coordinate for GWTC-3-confident events, illustrating how an unbounded axis can be mapped into a finite interval while preserving causal ordering intuitions. Data: GWOSC [5].

the statement being tested is not only local curvature. It is the global boundary condition that the exterior region is a black hole spacetime with a causal horizon, so that perturbations are absorbed rather than reflected back from a timelike wall.

Phenomenon 3: Horizon scale imaging is a direct probe of photon capture and causal separation. EHT images show a bright ring surrounding a dim interior for M87* and Sgr A* [9, 10]. The dim interior is not darkness because matter is absent; it is a consequence of photon capture by the horizon and extreme lensing near the photon orbit. The comparison in Figure 1.14 connects observed ring diameters to a GR shadow scale derived from mass and distance. The relevant global idea is that some null geodesics reach the observer while others terminate on the horizon. A Penrose diagram is the causality map that tells you which null directions belong to which class.

Phenomenon 4: Relativistic redshift in S2 is a measured slice of a causal diagram. Near pericenter, the S2 orbit is close enough to Sgr A* that gravitational redshift and transverse Doppler shift become detectable [11]. Figure 1.15 shows that the sum of these contributions lands near the reported observed scale. The observation does not directly draw a Penrose diagram, but it confirms the time dilation behavior that is consistent with a black hole exterior geometry. In global terms, it strengthens the empirical case that the Galactic center object is described by a causal structure that includes a horizon, even though the horizon itself is not locally detectable.

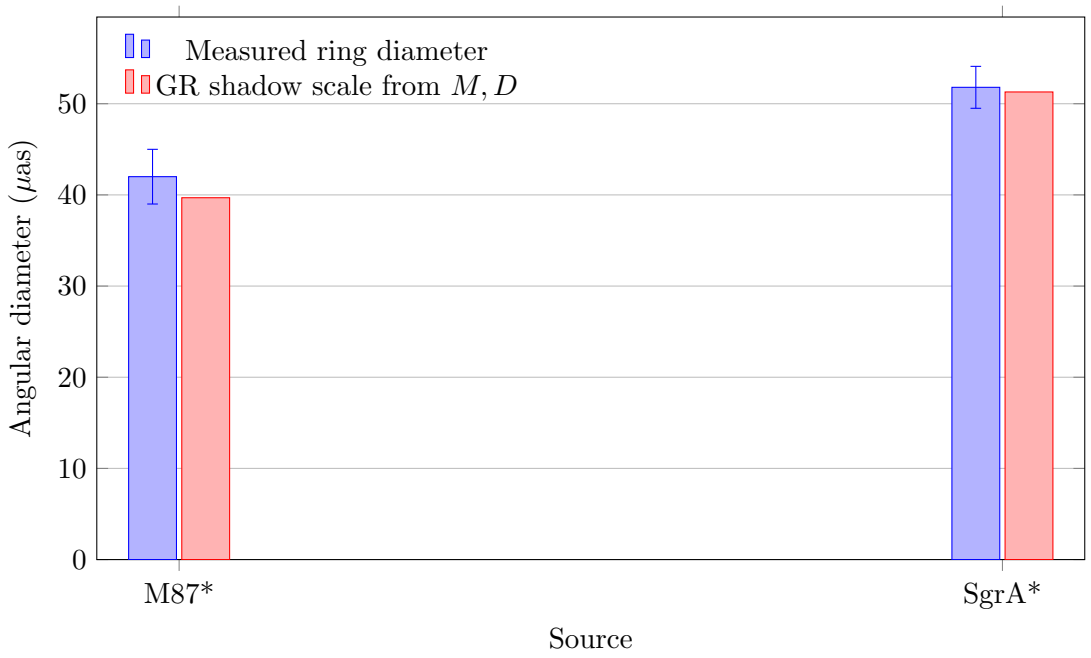


Figure 1.14: EHT ring diameters for M87* and Sgr A* compared with a GR shadow scale computed from the reported mass and distance. Measurements and reported values: [9, 10].

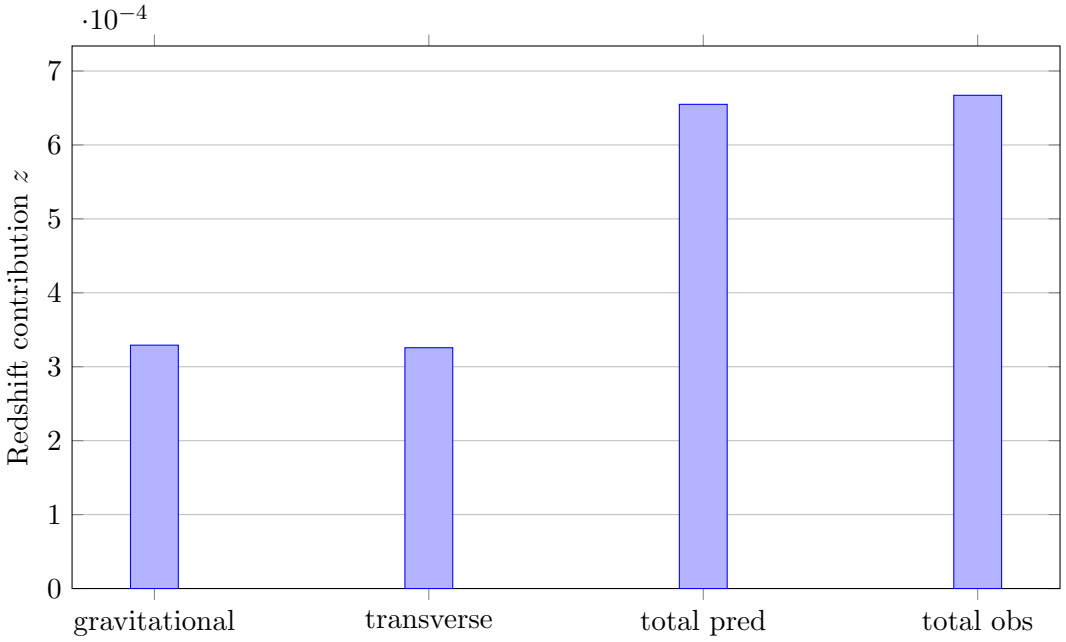


Figure 1.15: Redshift budget at S2 pericenter near Sgr A*: gravitational and transverse Doppler terms computed from reported pericenter distance and speed, and compared to the reported observed scale. Source parameters and observed redshift scale: [11].

Phenomenon 5: Trapped regions and predictability shape transient astrophysics. In dynamical collapse or merger, apparent horizons and trapped surfaces are quasi local di-

agnostics, while the event horizon is a global boundary. Observationally, transients such as tidal disruption events and rapid flares near Sgr A* are interpreted using models that assume a trapped region exists so that inflowing matter can disappear from causal contact with distant observers. The distinction matters when interpreting emission cutoffs: an apparent horizon can form and grow before the event horizon is fully determined, and the causal boundary that defines escape to infinity depends on the full future development.

1.8.2 A teleology sketch (conceptual)

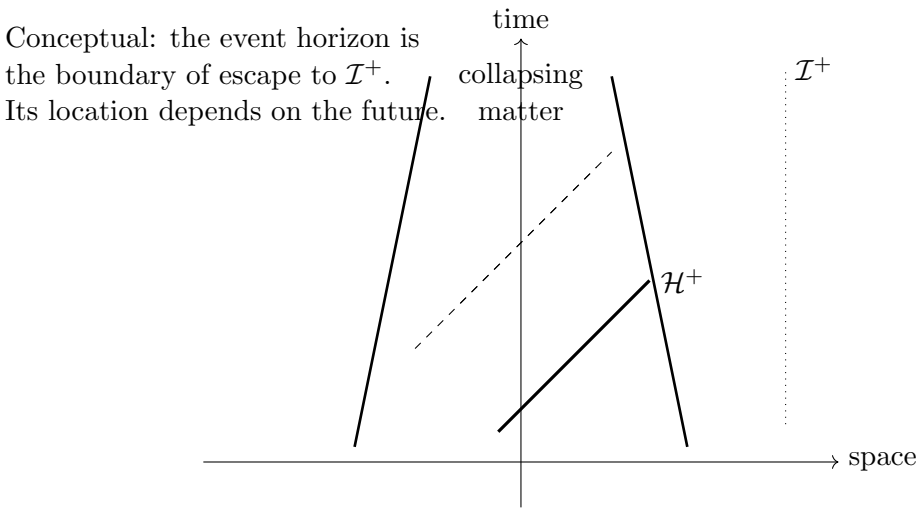


Figure 1.16: Conceptual diagram (not data derived). Teleological character of the event horizon: $\mathcal{H}^+$ is defined relative to $\mathcal{I}^+$ and cannot be located from local measurements alone [3].

1.9 Conclusion

Penrose diagrams are not decorative. They are the minimal language that keeps global causal statements from being smuggled into local coordinate artifacts. Conformal compactification makes infinity into a boundary, allowing clear definitions of $\mathcal{I}^{\pm}$ and making the event horizon a boundary of reachability rather than a surface of large curvature. In Schwarzschild, the horizon is a null surface generated by light rays that barely fail to escape, and the Kruskal extension shows that the full causal story includes regions that are hidden in standard coordinates [2].

The data driven plots in this chapter show how global concepts are used in observational practice. The GWTC-3-confident catalog provides masses, distances, and redshifts that can be turned into horizon radii, surface gravities, and light crossing times (Figures 1.5 through 1.12) [5]. Dynamical black hole masses in X-ray binaries provide an independent mass scale check and reinforce the linear mapping between measured mass and horizon radius (Figure 1.9) [8]. EHT ring diameters make photon capture scales visible in angle on the sky (Figure 1.14) [9, 10]. The S2 redshift detection provides a direct time dilation measurement near Sgr A* (Figure 1.15) [11].

Together, these datasets demonstrate a practical theme: “causal boundaries” are not only theory. They shape what is measured, how it is modeled, and what counts as an observational test.

Bibliography

[1] R. Penrose, "Conformal treatment of infinity," *General Relativity and Gravitation*, vol. 43, pp. 901–922, 2011 (republication of 1964 work). [Online]. Available: https://link.springer.com/content/pdf/10.1007/s10714-010-1110-5.pdf

[2] M. D. Kruskal, "Maximal extension of Schwarzschild metric," *Physical Review*, vol. 119, no. 5, pp. 1743–1745, 1960. [Online]. Available: https://harvest.aps.org/v2/journals/articles/10.1103/PhysRev.119.1743/fulltext

[3] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*. Cambridge University Press, 1973. [Online]. Available: https://www.cambridge.org/core/books/large-scale-structure-of-spacetime/1E6B961EC9878EDDBBD6AC0AF031CC93

[4] B. Carter, "Global structure of the Kerr family of gravitational fields," *Physical Review*, vol. 174, no. 5, pp. 1559–1571, 1968. [Online]. Available: https://people-lux.obspm.fr/carter/trav/Carter68.pdf

[5] Gravitational Wave Open Science Center, "GWTC-3-confident (ASCII export)," open event catalog data. [Online]. Available: https://gwosc.org/eventapi/ascii/GWTC-3-confident/

[6] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), "GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second half of the third observing run," 2021. [Online]. Available: https://arxiv.org/abs/2111.03606

[7] Gravitational Wave Open Science Center, "GWOSC API documentation," read only public endpoints for catalogs and event metadata. [Online]. Available: https://gwosc.org/api/

[8] G. Wiktorowicz and C. Belczynski (compiled table; hosted dataset), "Properties of Galactic black hole X-ray binaries (table)," 2013 compilation. [Online]. Available: https://stellarcollapse.org/sites/default/files/table.pdf

[9] The Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope results. I. The shadow of the supermassive black hole," 2019. [Online]. Available: https://arxiv.org/abs/1906.11238

[10] The Event Horizon Telescope Collaboration, "First Sagittarius A* Event Horizon Telescope results. I. The shadow of the supermassive black hole," 2023. [Online]. Available: https://arxiv.org/abs/2311.08680

[11] GRAVITY Collaboration (R. Abuter *et al.*), "Detection of the gravitational redshift in the orbit of the star S2 near the Galactic centre massive black hole," 2018. [Online]. Available: https://arxiv.org/abs/1807.09409

[12] J. M. Corral-Santana *et al.*, "BlackCAT: A catalogue of stellar-mass black holes in X-ray transients," 2015. [Online]. Available: https://arxiv.org/abs/1510.08869

Black Holes Geometry: Curvature and Horizons

Contents

II Einstein Geometry and Exact Black Hole Spacetimes **1**

8 Kerr Geometry: Rotation Warps Reality in a New Dimension **2**

Abstract . . . 2

8.1 Introduction . . . 3

8.2 Stationarity vs Staticity and Axial Symmetry . . . 4

8.3 Boyer Lindquist Coordinates and the Kerr Metric . . . 5

8.4 Horizons, the Ergosphere, and the Causal Meaning of Rotation . . . 5

8.5 Frame Dragging, ZAMOs, and Horizon Rotation . . . 7

8.6 Ring Singularity and Causal Oddities . . . 8

8.7 Geodesic Integrability and the Carter Constant . . . 9

8.8 Photon Regions and Shadow Geometry . . . 9

8.9 Data Driven Kerr Geometry from Gravitational Wave Catalogs . . . 11

8.9.1 Open dataset and methodology . . . 11

8.9.2 Dataset used for figures (inline tables) . . . 12

8.9.3 Data Figure 1: Component mass plane . . . 14

8.9.4 Data Figure 2: Mass ratio distribution . . . 14

8.9.5 Data Figure 3: Effective inspiral spin distribution . . . 14

8.9.6 Data Figure 4: Remnant Kerr spin distribution . . . 14

8.9.7 Data Figure 5: Remnant spin versus mass ratio . . . 14

8.9.8 Data Figure 6: Outer horizon radius ratio . . . 14

8.9.9 Data Figure 7: Surface gravity (dimensionless) distribution . . . 14

8.9.10 Data Figure 8: Horizon rotation frequency distribution . . . 14

8.9.11 Data Figure 9: ISCO frequency versus remnant mass . . . 14

8.9.12 Data Figure 10: Shadow ring diameters from EHT measurements 14

8.9.13 Interpretation and geometric takeaways . . . 16

8.10 Real World Applicability: Phenomenon Statements . . . 19

8.10.1 Phenomenon statement 1: Relativistic jets as a geometric energy extraction channel . . . 19

8.10.2 Phenomenon statement 2: Quasi-periodic oscillations and Lense Thirring precession . . . 19
8.10.3 Phenomenon statement 3: Gravitational wave ringdown as spectroscopy of Kerr curvature . . . 20
8.10.4 Phenomenon statement 4: Horizon scale imaging links photon regions to data . . . 20
8.10.5 Phenomenon statement 5: Accretion driven spin evolution and the approach to a spin limit . . . 20
8.10.6 Phenomenon statement 6: Spin aligned versus misaligned populations encode formation channels . . . 21
8.11 Conclusion . . . 21

List of Figures

8.1 Data based component mass plane for confident compact binary merger events (subset described in the text), using preferred parameter estimation source-frame masses from GWOSC [7, 8]. The concentration toward comparable masses reflects the detection bias toward higher chirp masses and the astrophysical distribution of binaries. 13

8.2 Data based histogram of mass ratio q for the catalog subset. The median $q \approx 0.653$ indicates that many detected mergers involve moderately comparable masses, which tends to increase remnant spin and radiated energy compared to highly asymmetric mergers. Data: GWOSC event portal snapshot [8]. 13

8.3 Data based histogram of the effective inspiral spin parameter χ_{eff}, a mass-weighted projection of component spins along the orbital angular momentum. Values near zero are common, while a tail to positive values indicates a population of binaries with partially aligned spins. Data: GWOSC [7, 8]. 14

8.4 Data based histogram of inferred remnant dimensionless Kerr spin a_f. The distribution clusters around $a_f \sim 0.693$, consistent with the expectation that mergers of comparable masses generically leave rapidly rotating remnants. Data: GWOSC [7, 8]. 15

8.5 Data based relationship between mass ratio and inferred remnant spin. In this subset the Pearson correlation between a_f and q is approximately 0.515, illustrating that more equal-mass mergers tend to leave faster-spinning Kerr remnants. Data: GWOSC [7, 8]. 15

8.6 Data based histogram of the dimensionless outer horizon radius ratio $r_+/M = 1 + \sqrt{1 - a_f^2}$ computed from inferred remnant spins. Smaller r_+/M corresponds to higher spin. This illustrates an invariant geometric statement: spin changes the horizon location relative to mass in Kerr. Data: GWOSC [7, 8]. 16

8.7 Data based histogram of the dimensionless surface gravity κM. In Kerr, κ decreases as spin approaches extremality, reflecting the geometric fact that extremal horizons have vanishing surface gravity. The plotted values are computed from inferred remnant spins using Kerr relations. Data: GWOSC [7, 8]. 16
8.8 Data based histogram of the horizon rotation frequency $f_H = \Omega_H/(2\pi)$ in Hz, computed from inferred remnant masses and spins using Eq. (8.8). For stellar-mass remnants, characteristic frequencies are tens to hundreds of Hz, linking Kerr horizon geometry to dynamical timescales seen in strong-field phenomena. Data: GWOSC [7, 8]. 17
8.9 Data based scatter of the prograde ISCO frequency $f_{\rm ISCO}$ computed from inferred remnant mass and spin. The inverse mass scaling is visible, while spin moves the ISCO inward and increases $f_{\rm ISCO}$ compared to the nonrotating case. Data: GWOSC [7, 8]. 17
8.10 Data based comparison of EHT reported ring diameters for M87* and Sgr A* with quoted uncertainties from the collaboration analyses [12, 13]. These angular scales are governed by null geodesic structure near the photon region in Kerr-like geometry. 18

List of Conceptual Diagrams

8.1 Conceptual comparison of static versus merely stationary geometry. In Kerr, stationarity persists but hypersurface orthogonality fails because rotation introduces twist, which is encoded by a $g_{t\phi}$ cross term in adapted coordinates. 4

8.2 Conceptual cross-section of the Kerr outer horizon r_+, inner horizon r_-, and the static limit (ergosurface) $r_E(\theta)$. The ergoregion sits between the static limit and the event horizon. 6

8.3 Conceptual diagram of the Penrose energy extraction process in the ergoregion. The geometry allows negative Killing energy states inside the ergoregion, enabling an escaping fragment to carry away more energy than the original infalling object. 7

8.4 Conceptual frame dragging: the local inertial directions are twisted around the rotating horizon. A ZAMO has zero axial angular momentum but still rotates relative to infinity. 8

8.5 Conceptual summary of Kerr integrability. The explicit symmetries give E and L_z, while a hidden symmetry gives the Carter constant Q, enabling separation into radial and polar potentials. 10

8.6 Conceptual depiction of the photon region and the origin of the black hole shadow as a capture set for null geodesics. In Kerr, rotation deforms the critical set by separating prograde and retrograde photon structures. . . . 10

Part II

Einstein Geometry and Exact Black Hole Spacetimes

Chapter 8

Kerr Geometry: Rotation Warps Reality in a New Dimension

Abstract

The Kerr solution is the canonical vacuum geometry of a rotating, astrophysically realistic black hole. Unlike Schwarzschild, Kerr is not merely a curved stage where matter moves, but a spacetime that actively mixes time and angle: rotation becomes geometry. In this chapter we build an intuition for Kerr as a precise Lorentzian manifold with axial symmetry, horizons, and an ergoregion where "rest" is not a coordinate choice but a physical impossibility. We then connect these geometric ideas to real measurements by treating public gravitational wave catalogs as a dataset of Kerr parameters. Using open access event summary values from the Gravitational Wave Open Science Center, we extract masses and remnant spins for confident compact binary mergers in the first three observing eras and compute derived horizon quantities (horizon radius, surface gravity, horizon angular velocity) and orbital landmarks (ISCO radius and ISCO frequency). The resulting data figures show how Kerr geometry is not an abstract artifact: it is a measurable object whose parameters are now routinely inferred from detector data, and whose qualitative features are directly related to phenomena such as frame dragging, precession, energy extraction, ringdown, and black hole shadows.

Keywords: Kerr spacetime, stationary axisymmetric geometry, Boyer Lindquist coordinates, event horizon, ergoregion, frame dragging, Carter constant, geodesic integrability, photon region, black hole shadow, gravitational wave catalogs, horizon surface gravity.

8.1 Introduction

Rotation is the rule rather than the exception in astrophysics. Stars rotate, disks rotate, galaxies rotate, and the compact remnants formed from gravitational collapse inherit angular momentum unless a miraculous cancellation occurs. In general relativity, angular momentum does not merely add a centrifugal term to Newtonian gravity; it reshapes the causal structure of the spacetime itself. The Kerr metric, discovered in 1963 by Roy Kerr [1], is the exact vacuum solution describing an asymptotically flat, stationary and axially symmetric black hole characterized only by its mass M and angular momentum J (equivalently a spin length scale $a = J/M$ in geometric units with $G = c = 1$). It is the unique endpoint expected for realistic rotating collapse in classical general relativity under broad assumptions, and it provides the geometric template behind many of the observables used in modern black hole astronomy.

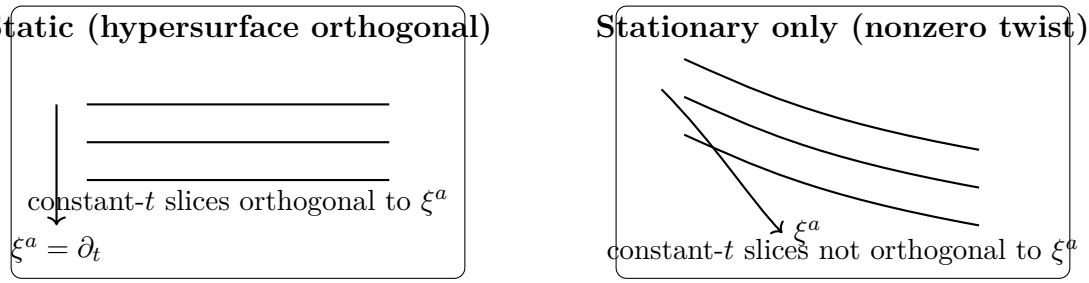
The conceptual jump from Schwarzschild to Kerr is not simply that things orbit differently or that the horizon is smaller. The deeper jump is that the spacetime now contains an intrinsic swirl: the time translation symmetry is no longer orthogonal to hypersurfaces of constant time. Equivalently, the timelike Killing field that represents stationarity has a nonzero twist, and the geometry forces local inertial frames to rotate relative to infinity. This is frame dragging, which in Kerr is not a perturbation but a global feature tied to the presence of the $g_{t\phi}$ cross term. In the language of differential geometry developed earlier in this book, Kerr is where symmetries become operational: Killing vectors are not only bookkeeping devices, they define conserved quantities along geodesics and allow separability of field equations [2, 3]. Kerr is also where horizon physics becomes richer. There are typically two horizons (outer and inner) and an ergoregion that supports energy extraction processes, including the classic Penrose process and the electromagnetic Blandford Znajek mechanism [5].

This chapter has two parallel goals. First, it builds geometric understanding: what stationary but not static means, why Boyer Lindquist coordinates are natural, how horizons and ergosurfaces arise from metric degeneracies, and why geodesics are integrable. Second, it turns that understanding into a data habit. Kerr is parameterized by (M, a), and gravitational wave astronomy now infers these parameters for merger remnants routinely. Therefore, Kerr geometry is no longer only a textbook exercise. It is a model class continuously confronted with catalogs of observed events. We use the Gravitational Wave Open Science Center event portal and its Zenodo snapshots as an open dataset [7, 8] and build data figures that illustrate horizon and orbital geometry inferred from observations.

8.2 Stationarity vs Staticity and Axial Symmetry

A spacetime is *stationary* if it admits a timelike Killing vector field ξ^a in some region, representing time translation symmetry. It is *static* if, in addition, ξ^a is hypersurface orthogonal, meaning there exist spacelike hypersurfaces orthogonal to ξ^a everywhere in that region. Geometrically, hypersurface orthogonality is the Frobenius condition $\xi_{[a}\nabla_b\xi_{c]} = 0$, which can be interpreted as the vanishing of the twist of ξ^a. Schwarzschild is static; Kerr is stationary but not static. The difference is not semantic. In a static spacetime, there is a globally meaningful notion of observers at rest with respect to the symmetry, at least outside horizons. In Kerr, the spacetime itself carries angular momentum, and the would-be rest frames are forced to rotate.

Axial symmetry is encoded by another Killing vector ψ^a with closed orbits, typically identified with ∂_ϕ in adapted coordinates. Stationarity and axial symmetry together mean Kerr admits two commuting Killing fields, ∂_t and ∂_ϕ, which generate an $\mathbb{R} \times U(1)$ symmetry group. This symmetry is the geometric origin of two conserved quantities for geodesic motion: energy $E = -p_a(\partial_t)^a$ and azimuthal angular momentum $L_z = p_a(\partial_\phi)^a$. What makes Kerr special is that it admits an additional hidden symmetry, encoded in a rank two Killing tensor, giving the Carter constant Q [3]. The existence of Q is not a small technicality. It elevates Kerr from complicated to integrable and permits analytic understanding of many orbit families.



Conceptual Diagram 8.1: Conceptual comparison of static versus merely stationary geometry. In Kerr, stationarity persists but hypersurface orthogonality fails because rotation introduces twist, which is encoded by a $g_{t\phi}$ cross term in adapted coordinates.

The diagram above is conceptual, but it supports a precise geometric statement: if the metric admits coordinates adapted to the symmetries, then staticity corresponds to the possibility of eliminating all $dt\,d\phi$ mixing terms by coordinate choices that preserve time translation symmetry. Kerr does not allow that elimination globally outside the horizon. The cross term is not a coordinate annoyance; it is the tensorial imprint of the spacetime's angular momentum.

8.3 Boyer Lindquist Coordinates and the Kerr Metric

To write Kerr explicitly, one introduces coordinates (t, r, θ, ϕ) adapted to the symmetries (time translation and axial rotation). Boyer and Lindquist provided the standard form [2], which resembles Schwarzschild but with key modifications. Define

$$\Sigma = r^2 + a^2 \cos^2\theta, \qquad \Delta = r^2 - 2Mr + a^2. \tag{8.1}$$

In Boyer Lindquist coordinates, the Kerr line element is

$$\begin{aligned} ds^2 = & -\left(1 - \frac{2Mr}{\Sigma}\right) dt^2 - \frac{4Mar\sin^2\theta}{\Sigma}\, dt\, d\phi + \frac{\Sigma}{\Delta}\, dr^2 + \Sigma\, d\theta^2 \\ & + \left(r^2 + a^2 + \frac{2Ma^2 r \sin^2\theta}{\Sigma}\right) \sin^2\theta\, d\phi^2. \end{aligned} \tag{8.2}$$

Several geometric features are already visible. First, Δ controls the radial coefficient $g_{rr} = \Sigma/\Delta$, so roots of Δ signal coordinate degeneracies that will become horizons. Second, the mixed term $g_{t\phi}$ is proportional to a, vanishing in the nonrotating limit and encoding frame dragging when $a \neq 0$. Third, Σ controls the departure from spherical symmetry: surfaces of constant r are not perfect spheres but oblate spheroids, more squashed when a is larger.

This geometric fact can be stated precisely: r is not a naive Euclidean radius but a coordinate labeling of confocal quadrics in the Kerr geometry. A useful habit is to treat coordinates as labels tied to symmetry, not as literal rulers. When one computes invariant quantities such as curvature scalars or horizon area, the coordinate ambiguities disappear and geometry becomes operational.

8.4 Horizons, the Ergosphere, and the Causal Meaning of Rotation

In Kerr, horizons arise where $\Delta = 0$. Solving $\Delta = r^2 - 2Mr + a^2 = 0$ yields two radii

$$r_\pm = M \pm \sqrt{M^2 - a^2}, \tag{8.3}$$

provided $|a| \leq M$. The outer solution r_+ is the event horizon of the Kerr black hole in the maximally extended geometry. The inner solution r_- is a Cauchy horizon, which plays a role in global structure and inner horizon instability discussions. The equality case $|a| = M$ corresponds to extremality, where the two horizons coincide. If $|a| > M$, the roots disappear and the geometry contains a naked singularity, a situation widely believed

to be forbidden in nature by some form of cosmic censorship, although this remains a deep conjecture rather than a theorem.

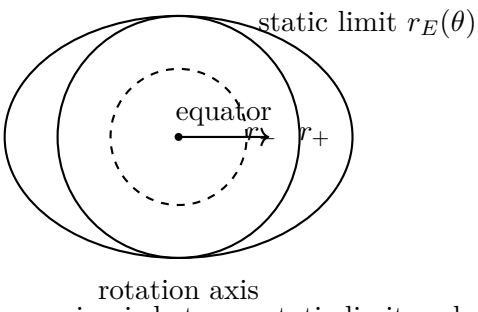
The ergoregion is defined not by Δ but by the sign change of g_{tt}. From (8.2),

$$g_{tt} = -\left(1 - \frac{2Mr}{\Sigma}\right). \tag{8.4}$$

The *static limit* occurs where $g_{tt} = 0$, giving

$$r_E(\theta) = M + \sqrt{M^2 - a^2\cos^2\theta}. \tag{8.5}$$

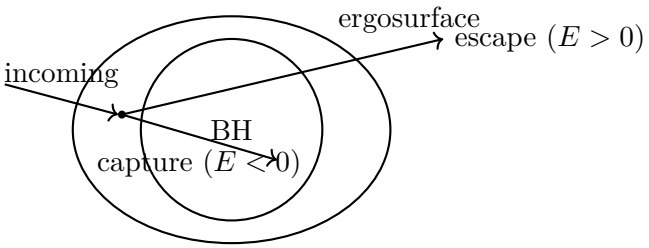
Outside $r_E(\theta)$, ∂_t is timelike and observers can remain at fixed (r, θ, ϕ). Inside $r_E(\theta)$ but outside r_+, ∂_t becomes spacelike. This does not mean time becomes space; it means that an observer following ∂_t would need to move faster than light. No physical observer can remain at rest relative to infinity. All timelike worldlines must acquire a nonzero ϕ-component. This is the most geometric and least metaphorical definition of frame dragging: the set of allowed timelike directions in the tangent space tilts so strongly that standing still is excluded.



Cross-section sketch: ergoregion is between static limit and outer horizon.

Conceptual Diagram 8.2: Conceptual cross-section of the Kerr outer horizon r_+, inner horizon r_-, and the static limit (ergosurface) $r_E(\theta)$. The ergoregion sits between the static limit and the event horizon.

A remarkable physical consequence follows immediately: within the ergoregion, the conserved energy associated with ∂_t can be negative for certain particle trajectories. This makes energy extraction possible. In the Penrose process, a particle enters the ergoregion and splits into two fragments; one falls into the hole on a negative-energy trajectory, while the other escapes to infinity with energy exceeding that of the original particle. This is not a perpetual motion machine. The extra energy comes from the black hole's rotational energy, reducing its spin. In wave language, the same geometric possibility underlies superradiant scattering, where certain modes extract rotational energy.



Penrose process: negative-energy capture inside ergoregion allows energy extraction.

Conceptual Diagram 8.3: Conceptual diagram of the Penrose energy extraction process in the ergoregion. The geometry allows negative Killing energy states inside the ergoregion, enabling an escaping fragment to carry away more energy than the original infalling object.

8.5 Frame Dragging, ZAMOs, and Horizon Rotation

One of the cleanest geometric ways to describe frame dragging is to ask: what angular velocity must an observer have in order to have zero angular momentum about the symmetry axis? Such observers are called ZAMOs (zero angular momentum observers). Because Kerr admits the axial Killing field ∂_ϕ, angular momentum about the axis is the conserved quantity

$$L_z = p_\phi. \tag{8.6}$$

A ZAMO has $L_z = 0$ but generally nonzero $d\phi/dt$. In Boyer Lindquist coordinates the ZAMO angular velocity is

$$\omega(r, \theta) = -\frac{g_{t\phi}}{g_{\phi\phi}}, \tag{8.7}$$

which is nonzero whenever $a \neq 0$. Thus, even if you arrange to have zero axial angular momentum, the geometry still makes you rotate relative to infinity. This is frame dragging as a statement about the structure of conserved quantities. It is also frame dragging as a statement about local inertial frames: free-fall gyroscopes precess, orbital planes precess, and accretion flows can precess when tilted.

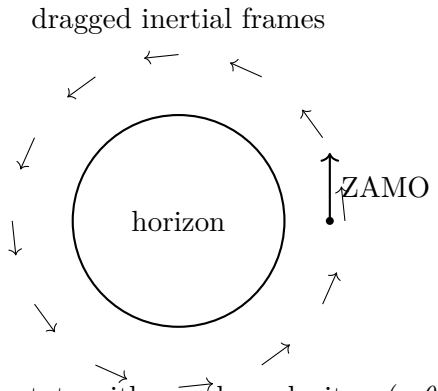
A special value is the horizon angular velocity, the angular velocity of the null generators of the event horizon. In geometric units it can be written as

$$\Omega_H = \frac{a}{r_+^2 + a^2} = \frac{a}{2Mr_+}. \tag{8.8}$$

It is often useful to express this in terms of a dimensionless spin parameter $a_* = a/M$, so that

$$\Omega_H M = \frac{a_*}{2\left(1 + \sqrt{1 - a_*^2}\right)}. \tag{8.9}$$

This formula will reappear in our data analysis section, where a_* is inferred for merger remnants and Ω_H becomes an inferred rotational frequency of the horizon itself.



Conceptual view: ZAMOs rotate with angular velocity $\omega(r,\theta)$ even when $L_z = 0$.

Conceptual Diagram 8.4: Conceptual frame dragging: the local inertial directions are twisted around the rotating horizon. A ZAMO has zero axial angular momentum but still rotates relative to infinity.

8.6 Ring Singularity and Causal Oddities

Curvature singularities in Kerr are not located at a point $r = 0$ as in Schwarzschild. Instead, the true singularity is where $\Sigma = r^2 + a^2\cos^2\theta = 0$, which occurs only when $r = 0$ and $\theta = \pi/2$. This is a ring singularity in the equatorial plane. The difference matters because the Kerr interior is not merely a smaller copy of Schwarzschild. In the maximally extended analytic continuation, one encounters regions where the coordinate r becomes negative and where closed timelike curves appear. These features are often discussed to illustrate the richness and potential pathologies of classical solutions. From a physical perspective, however, it is widely expected that the inner region is unstable and that realistic collapse, perturbations, and quantum effects modify the idealized interior. Nevertheless, the ring singularity and the existence of an inner horizon are instructive because they show that even vacuum curvature constraints can yield global causal structure far more intricate than naive intuition suggests.

A practical geometric lesson is that horizons and singularities are distinct concepts. The event horizon is a null hypersurface defined globally (it is the boundary of the causal past of future null infinity). The singularity is the breakdown of the spacetime manifold, detectable by divergence of curvature invariants or by geodesic incompleteness. In Kerr, the outer horizon r_+ is a regular null hypersurface for $|a| < M$, while the ring singularity is hidden behind it (in the idealized classical geometry). Thus rotation does not eliminate the horizon; it changes its geometry and adds the ergoregion.

8.7 Geodesic Integrability and the Carter Constant

A hallmark of Kerr is that the geodesic equations admit separation of variables. This is not automatic in a generic stationary axisymmetric spacetime, and it reflects the hidden symmetry found by Carter [3]. The starting point is the Hamilton Jacobi equation for geodesics:

$$g^{ab}\frac{\partial S}{\partial x^a}\frac{\partial S}{\partial x^b} = -\mu^2, \tag{8.10}$$

where μ is the rest mass (zero for photons) and S is Hamilton's principal function. In Kerr, one can take an ansatz of the form

$$S = -Et + L_z\phi + S_r(r) + S_\theta(\theta), \tag{8.11}$$

and the equation separates into a radial part and an angular part, introducing a separation constant Q (the Carter constant). The resulting first integrals of motion can be written as

$$\Sigma^2\left(\frac{dr}{d\lambda}\right)^2 = R(r), \tag{8.12}$$

$$\Sigma^2\left(\frac{d\theta}{d\lambda}\right)^2 = \Theta(\theta), \tag{8.13}$$

with

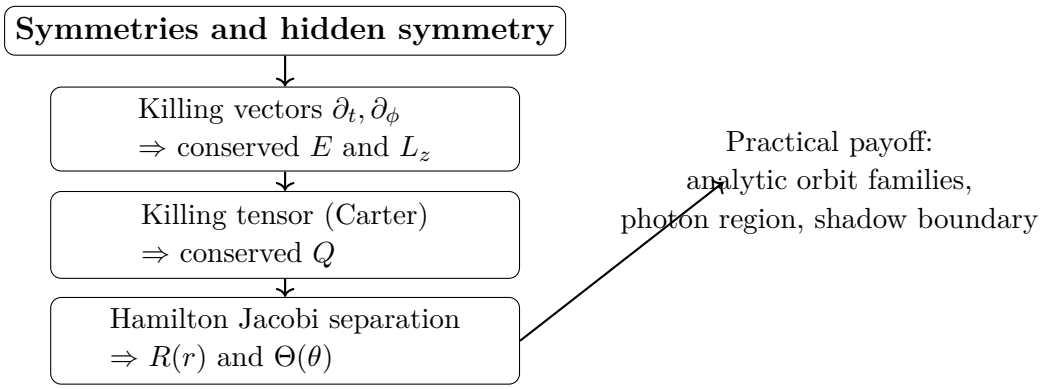
$$R(r) = \left[E(r^2+a^2) - aL_z\right]^2 - \Delta\left[\mu^2 r^2 + (L_z - aE)^2 + Q\right], \tag{8.14}$$

$$\Theta(\theta) = Q - \cos^2\theta\left[a^2(\mu^2 - E^2) + \frac{L_z^2}{\sin^2\theta}\right]. \tag{8.15}$$

The structure is revealing. The motion is encoded in two effective potentials $R(r)$ and $\Theta(\theta)$, and the allowed region is where both are nonnegative. The same constants (E, L_z, Q) also control photon trajectories, and therefore lensing, photon regions, and shadows. This is why the Carter constant, though abstract, becomes observationally relevant: it organizes null geodesics, which organize images.

8.8 Photon Regions and Shadow Geometry

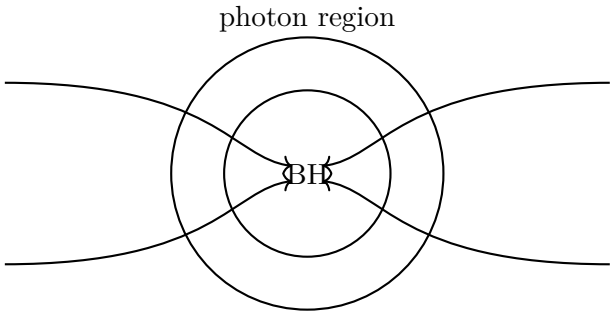
The black hole shadow is the set of directions on an observer's sky that correspond to null geodesics which fall into the horizon rather than reaching distant sources. In Kerr, the shadow is not perfectly circular for generic observers; its shape depends on spin and inclination, and is controlled by the structure of unstable photon orbits. A key concept is the photon region: the set of bound null geodesics that remain at constant



Conceptual Diagram 8.5: Conceptual summary of Kerr integrability. The explicit symmetries give E and L_z, while a hidden symmetry gives the Carter constant Q, enabling separation into radial and polar potentials.

radius (spherical photon orbits). These orbits satisfy conditions $R(r) = 0$ and $dR/dr = 0$ for photons ($\mu = 0$). Solving these conditions yields critical impact parameters, which translate into the shadow edge on the observer's image plane.

Although the full mapping from (E, L_z, Q) to observed angles involves a choice of tetrad for the observer, the qualitative story is robust: rotation breaks the symmetry between prograde and retrograde photon trajectories. Photons co-rotating with the hole can orbit closer, while counter-rotating photons orbit farther. This asymmetry shifts the shadow center and distorts its shape. In the slow rotation limit, the distortion is small, but in near-extremal Kerr it becomes pronounced. The observational significance is that imaging experiments such as the Event Horizon Telescope (EHT) constrain combinations of mass, spin, and inclination by measuring the size and asymmetry of the emission ring around the shadow [12, 13].



Conceptual lensing: some null geodesics scatter, others plunge. The capture set defines the shadow.

Conceptual Diagram 8.6: Conceptual depiction of the photon region and the origin of the black hole shadow as a capture set for null geodesics. In Kerr, rotation deforms the critical set by separating prograde and retrograde photon structures.

8.9 Data Driven Kerr Geometry from Gravitational Wave Catalogs

8.9.1 Open dataset and methodology

Modern gravitational wave catalogs provide posterior summaries for compact binary coalescences detected by LIGO, Virgo, and KAGRA. These events are among the cleanest available probes of strong-field Kerr geometry because the late time merger and ringdown are governed primarily by the mass and spin of the remnant black hole, which is expected to settle to a Kerr state on dynamical timescales. In this section we use the event metadata released through the Gravitational Wave Open Science Center (GWOSC) event portal [7]. To ensure reproducibility without proprietary access, we specifically rely on the Zenodo snapshot of the GWOSC event portal dated 2025-08-27 [8], which contains JSON metadata for the cumulative gravitational wave transient catalog (GWTC) and individual event pages.

Selection. From the snapshot we select the confident detections from GWTC-1-confident [9], GWTC-2.1-confident [10], and GWTC-3-confident [11]. For each event we use the preferred parameter estimation summary provided in the event metadata. We then exclude GW170817 because its merger remnant is not reliably a Kerr black hole on observational timescales, and our horizon formulas assume a black hole remnant. This yields $N = 88$ merger remnants with inferred remnant mass M_f and dimensionless remnant spin a_f.

Note on uncertainty. The event portal JSON provides posterior summaries (typically medians and credible intervals) rather than raw posterior samples. For compactness and reproducibility inside a single LaTeX file, the figures below use the preferred point estimates as representative values. All derived Kerr quantities are therefore illustrative distributions of inferred central values, not full uncertainty-propagated posteriors.

Derived geometric quantities. The data include $(M_f,\, a_f)$ as inference products. To connect to geometry, we compute the following Kerr quantities in units with $G = c = 1$:

$$r_\pm/M = 1 \pm \sqrt{1 - a_f^2}, \tag{8.16}$$

$$\frac{A}{16\pi M^2} = \frac{1}{2}\left(1 + \sqrt{1 - a_f^2}\right), \tag{8.17}$$

$$\kappa M = \frac{\sqrt{1 - a_f^2}}{2\left(1 + \sqrt{1 - a_f^2}\right)}, \tag{8.18}$$

$$\Omega_H M = \frac{a_f}{2\left(1 + \sqrt{1 - a_f^2}\right)}. \tag{8.19}$$

Here A is horizon area and κ is surface gravity. For orbital geometry we compute the prograde equatorial ISCO radius $r_{\rm ISCO}/M$ using the standard analytic expression [4] and convert it to a physical ISCO frequency in Hz for a test particle orbiting a Kerr black hole with mass M_f:

$$f_{\rm ISCO} = \frac{1}{2\pi M_{\rm geo}\left((r_{\rm ISCO}/M)^{3/2} + a_f\right)}, \tag{8.20}$$

where $M_{\rm geo} = (GM_f/c^3)$ is the mass expressed as a time, with $GM_\odot/c^3 \approx 4.925\,490\,947 \times 10^{-6}\,\mathrm{s}$.

Summary statistics. Table 8.1 gives representative percentiles for several inferred and derived quantities. A key qualitative result is that inferred remnant spins cluster around $a_f \sim 0.7$. This is not arbitrary. In vacuum general relativity, angular momentum balance in mergers of comparable masses naturally yields remnants with substantial spin, and accretion processes in thin disks are also expected to spin holes up toward a limiting value below extremality (the Thorne limit) [6].

Table 8.1: Percentiles (10th, 50th, 90th) for inferred and derived Kerr quantities computed from the GWOSC catalog subset described in the text.

Quantity	10th percentile	Median	90th percentile
a_f (dimensionless remnant spin)	0.529	0.693	0.811
$\chi_{\rm eff}$ (effective inspiral spin)	-0.128	0.046	0.281
M_f ($M_\odot$)	18.82	49.33	84.81
$q = m_2/m_1$	0.415	0.653	0.94
r_+/M	1.585	1.721	1.849
κM	0.165	0.202	0.245
f_H (Hz)	15.177	61.048	178.448
$f_{\rm ISCO}$ (Hz)	20.261	75.202	231.906

8.9.2 Dataset used for figures (inline tables)

The following inline tables store the dataset and the derived quantities used in the data figures. All plots in Figures 8.1 to 8.10 are generated directly from these tabulated values.

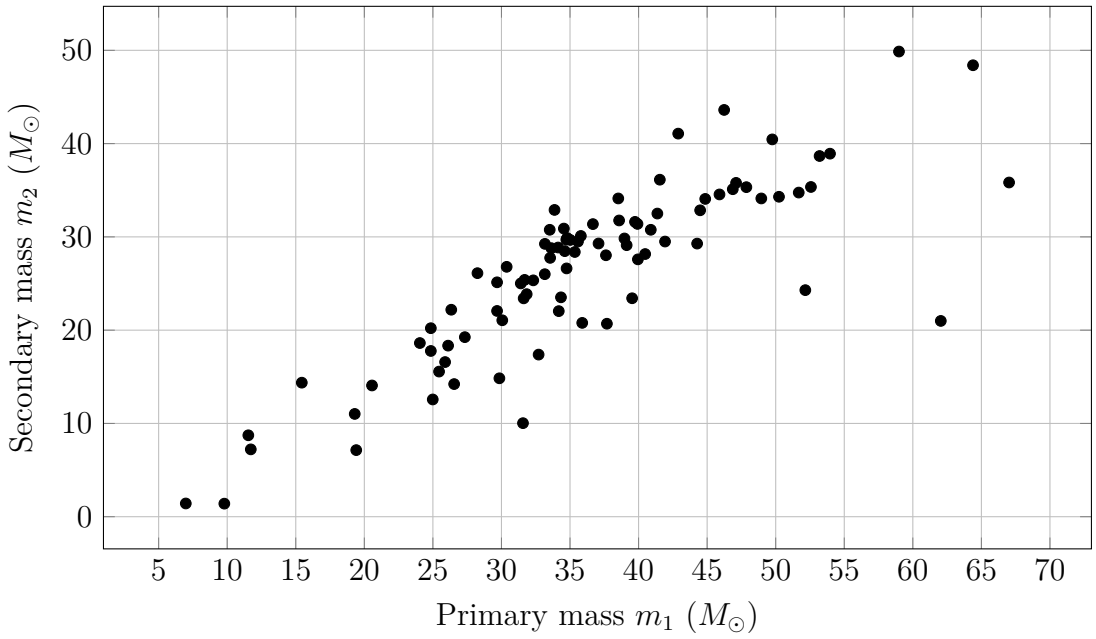


Figure 8.1: Data based component mass plane for confident compact binary merger events (subset described in the text), using preferred parameter estimation source-frame masses from GWOSC [7, 8]. The concentration toward comparable masses reflects the detection bias toward higher chirp masses and the astrophysical distribution of binaries.

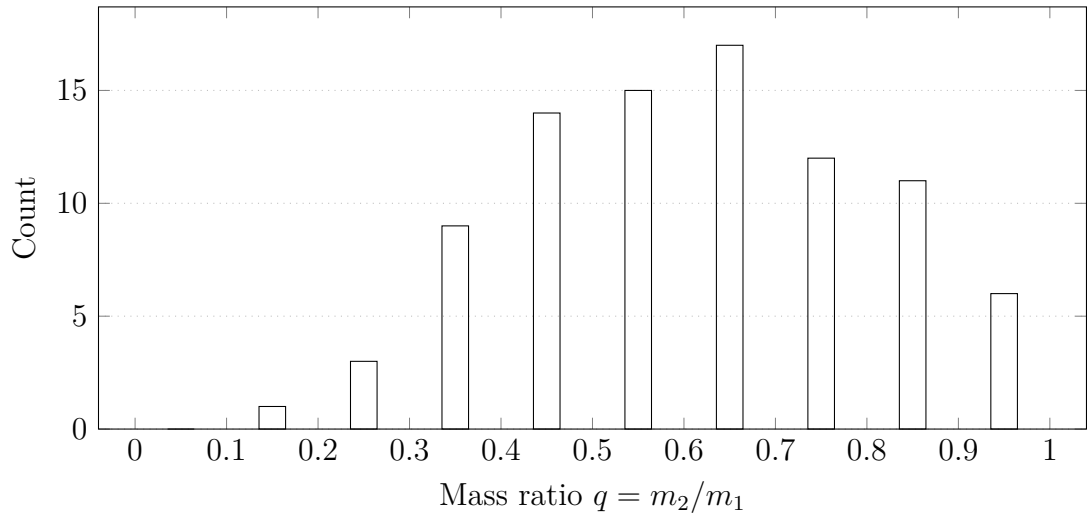


Figure 8.2: Data based histogram of mass ratio q for the catalog subset. The median $q \approx 0.653$ indicates that many detected mergers involve moderately comparable masses, which tends to increase remnant spin and radiated energy compared to highly asymmetric mergers. Data: GWOSC event portal snapshot [8].

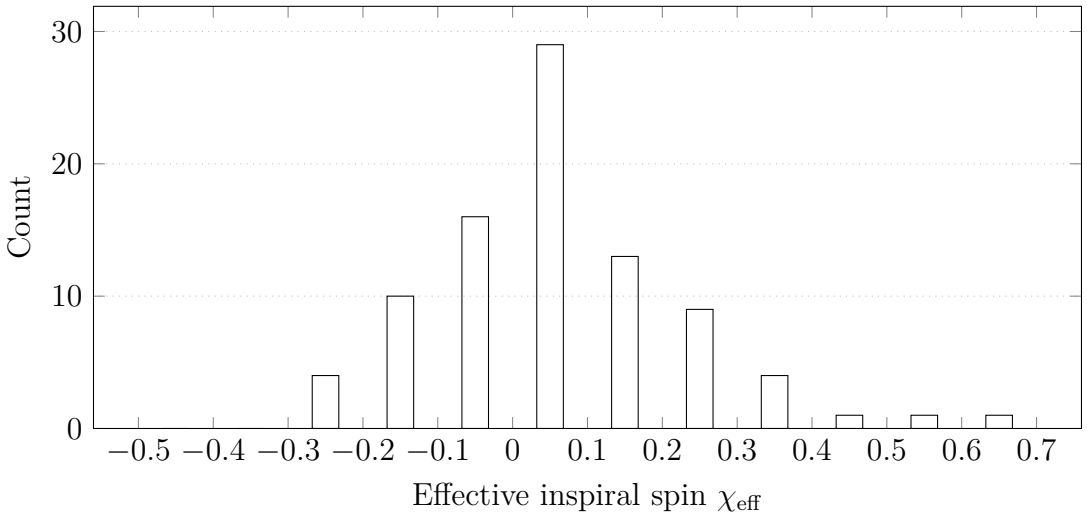


Figure 8.3: Data based histogram of the effective inspiral spin parameter χ_{eff}, a mass-weighted projection of component spins along the orbital angular momentum. Values near zero are common, while a tail to positive values indicates a population of binaries with partially aligned spins. Data: GWOSC [7, 8].

8.9.3 Data Figure 1: Component mass plane

8.9.4 Data Figure 2: Mass ratio distribution

8.9.5 Data Figure 3: Effective inspiral spin distribution

8.9.6 Data Figure 4: Remnant Kerr spin distribution

8.9.7 Data Figure 5: Remnant spin versus mass ratio

8.9.8 Data Figure 6: Outer horizon radius ratio

8.9.9 Data Figure 7: Surface gravity (dimensionless) distribution

8.9.10 Data Figure 8: Horizon rotation frequency distribution

8.9.11 Data Figure 9: ISCO frequency versus remnant mass

8.9.12 Data Figure 10: Shadow ring diameters from EHT measurements

The Kerr photon region and the associated shadow boundary underlie the approximate size of the bright emission ring seen in horizon-scale imaging. The EHT collaboration reports characteristic ring diameters for M87* and Sgr A* in their imaging analyses

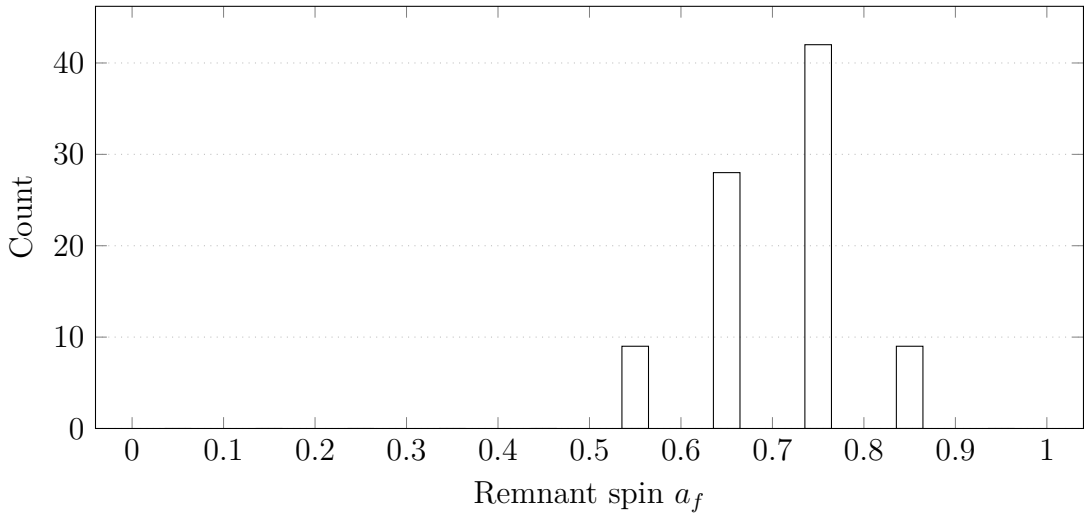


Figure 8.4: Data based histogram of inferred remnant dimensionless Kerr spin a_f. The distribution clusters around $a_f \sim 0.693$, consistent with the expectation that mergers of comparable masses generically leave rapidly rotating remnants. Data: GWOSC [7, 8].

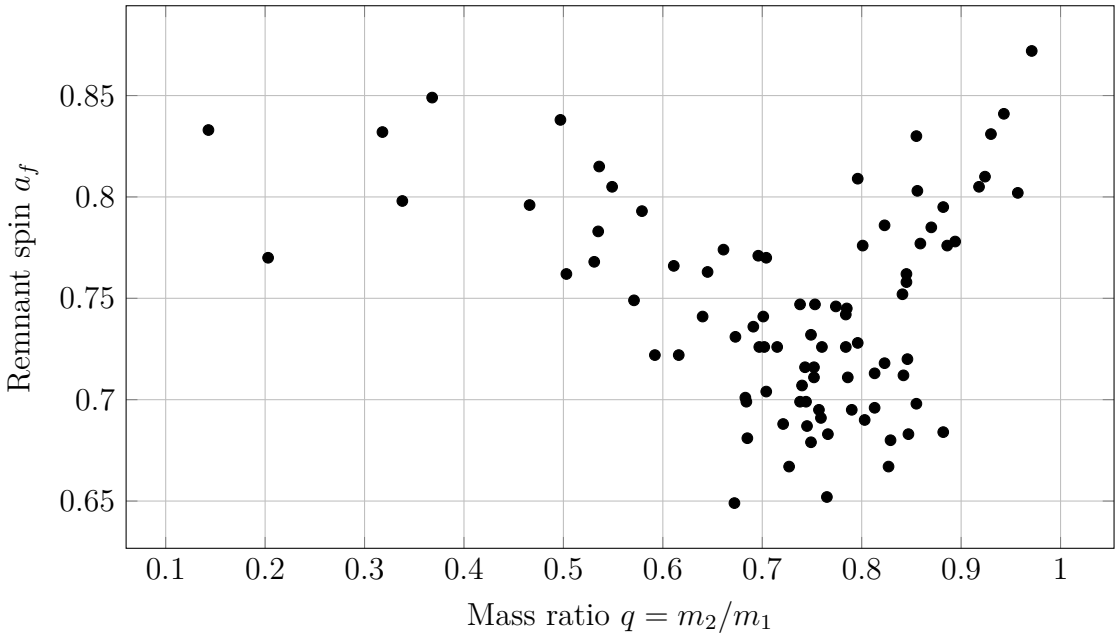


Figure 8.5: Data based relationship between mass ratio and inferred remnant spin. In this subset the Pearson correlation between a_f and q is approximately 0.515, illustrating that more equal-mass mergers tend to leave faster-spinning Kerr remnants. Data: GWOSC [7, 8].

[12, 13]. Although the ring diameter depends weakly on spin for many emission models, the measured angular scale provides a direct link between null geodesic geometry and

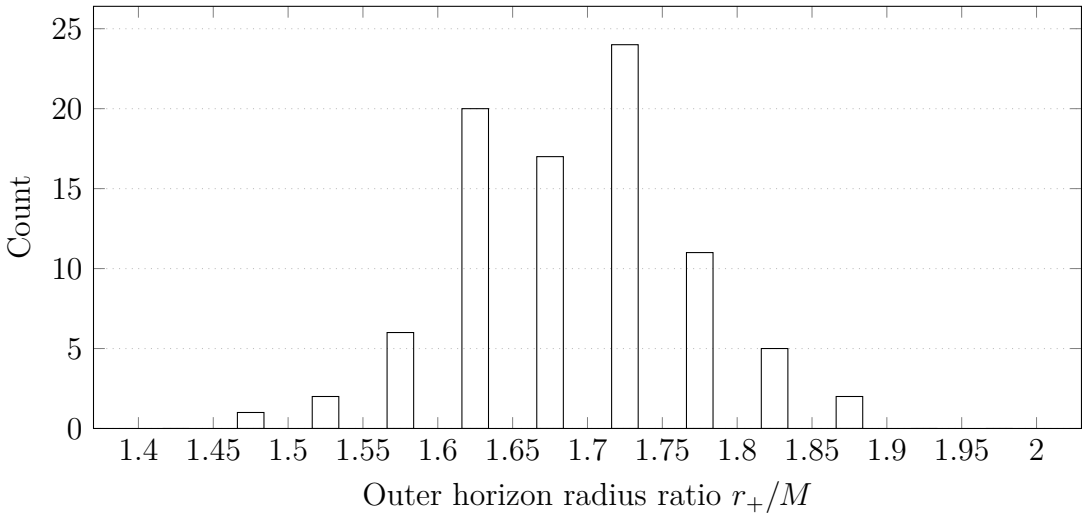


Figure 8.6: Data based histogram of the dimensionless outer horizon radius ratio $r_+/M = 1 + \sqrt{1 - a_f^2}$ computed from inferred remnant spins. Smaller r_+/M corresponds to higher spin. This illustrates an invariant geometric statement: spin changes the horizon location relative to mass in Kerr. Data: GWOSC [7, 8].

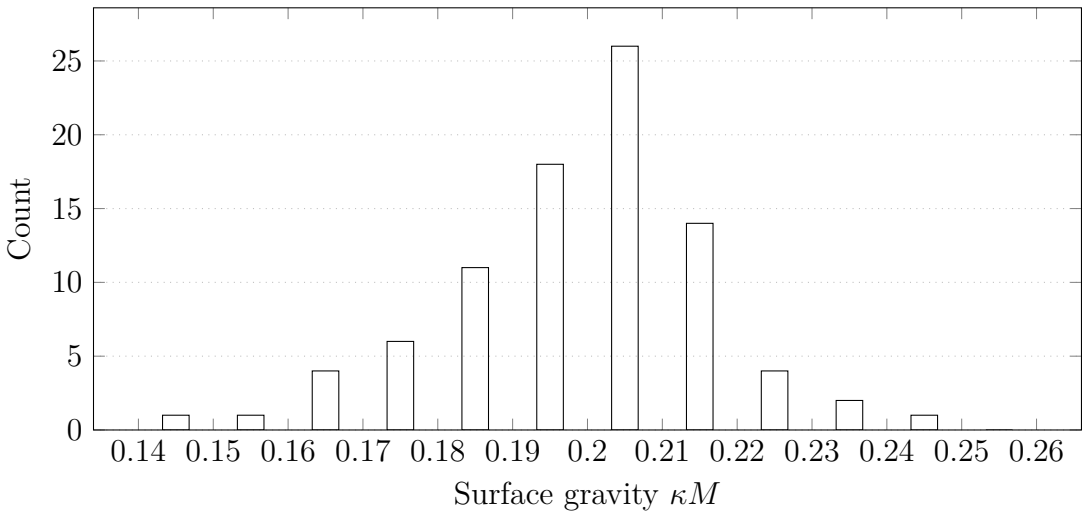


Figure 8.7: Data based histogram of the dimensionless surface gravity κM. In Kerr, κ decreases as spin approaches extremality, reflecting the geometric fact that extremal horizons have vanishing surface gravity. The plotted values are computed from inferred remnant spins using Kerr relations. Data: GWOSC [7, 8].

astronomical data.

8.9.13 Interpretation and geometric takeaways

Several geometric lessons emerge directly from the catalog plots.

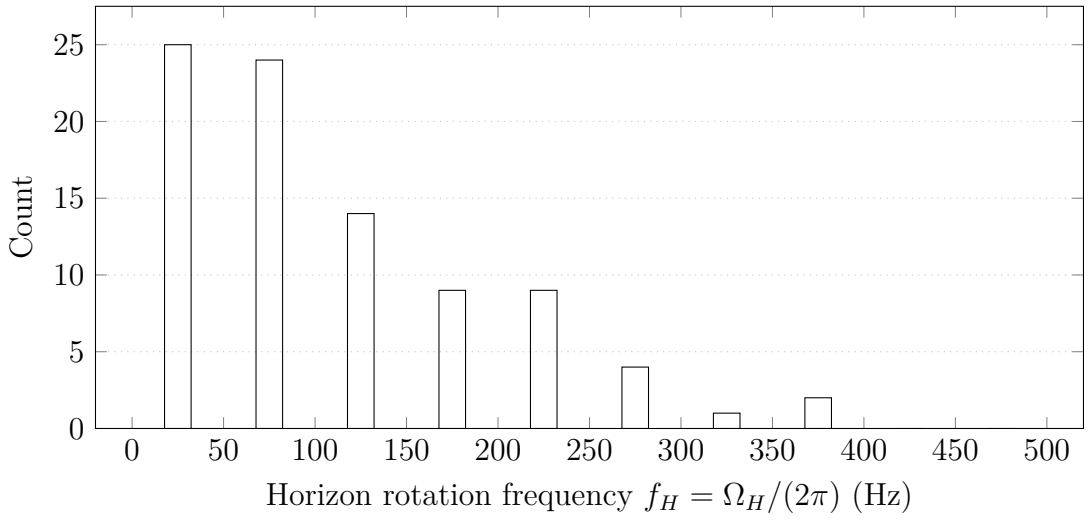


Figure 8.8: Data based histogram of the horizon rotation frequency $f_H = \Omega_H/(2\pi)$ in Hz, computed from inferred remnant masses and spins using Eq. (8.8). For stellar-mass remnants, characteristic frequencies are tens to hundreds of Hz, linking Kerr horizon geometry to dynamical timescales seen in strong-field phenomena. Data: GWOSC [7, 8].

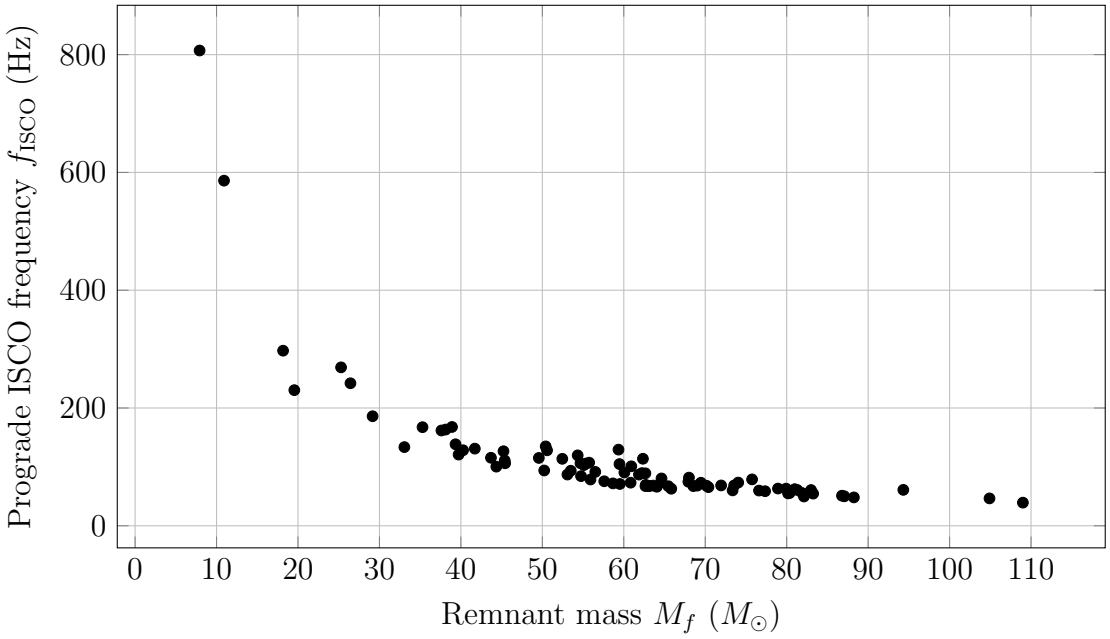


Figure 8.9: Data based scatter of the prograde ISCO frequency f_{ISCO} computed from inferred remnant mass and spin. The inverse mass scaling is visible, while spin moves the ISCO inward and increases f_{ISCO} compared to the nonrotating case. Data: GWOSC [7, 8].

First, the remnant spin distribution in Figure 8.4 being concentrated near $a_f \sim 0.7$ implies that the detected merger remnants are far from the nonrotating limit. Thus the

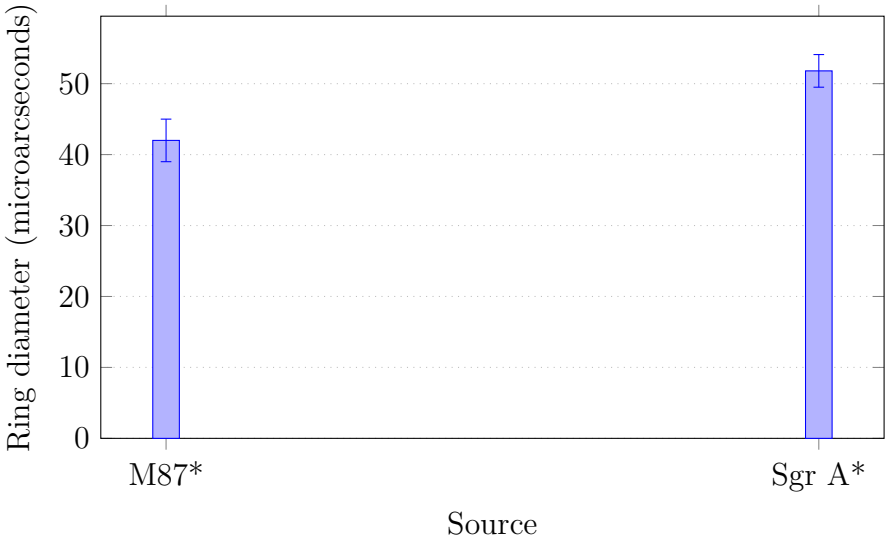


Figure 8.10: Data based comparison of EHT reported ring diameters for M87* and Sgr A* with quoted uncertainties from the collaboration analyses [12, 13]. These angular scales are governed by null geodesic structure near the photon region in Kerr-like geometry.

ergoregion is generically present, and the horizon radius ratio r_+/M is generically below 2 as shown in Figure 8.6. This means that many real black holes formed in mergers have a horizon that, in geometric units, is noticeably closer in than Schwarzschild. This is not a coordinate statement; it is invariantly encoded in the horizon area and the null generator properties.

Second, the dependence of a_f on mass ratio in Figure 8.5 illustrates how Kerr parameters are not arbitrary initial conditions but are dynamically produced. When q is close to unity, the binary carries substantial orbital angular momentum relative to its mass, and the remnant inherits a large fraction of it. This is why equal-mass mergers are efficient producers of rapidly rotating Kerr remnants. In contrast, when q is small, the orbital angular momentum per unit mass is reduced and the remnant spin decreases.

Third, surface gravity and horizon rotation are not abstract. They set characteristic timescales. The distribution of κM in Figure 8.7 indicates that these remnants are not near-extremal, so κ remains finite. The horizon frequency distribution in Figure 8.8 shows that for stellar-mass remnants, the horizon rotates on timescales of milliseconds to tens of milliseconds. These timescales are precisely in the band where gravitational wave ringdown and many accretion-driven variability phenomena occur, making the link between Kerr geometry and observation more than qualitative.

Finally, the ISCO frequency plot in Figure 8.9 shows how orbital geometry depends jointly on mass and spin. Because $f_{\mathrm{ISCO}} \propto 1/M$ at fixed dimensionless spin, higher-mass remnants have lower characteristic orbital frequencies. This scaling helps explain

why high-frequency timing signatures are expected from stellar-mass black holes, while supermassive black holes vary on much longer timescales. Kerr geometry thus provides a scaling map from mass and spin to observational timescales.

8.10 Real World Applicability: Phenomenon Statements

The central message of Kerr geometry is that rotation is geometry. Below are real-world phenomenon statements illustrating how the specific geometric structures discussed in this chapter become observable.

8.10.1 Phenomenon statement 1: Relativistic jets as a geometric energy extraction channel

Many accreting black holes launch collimated jets that carry enormous power. A leading theoretical framework relates jet power to electromagnetic extraction of rotational energy from a Kerr black hole, a process often modeled by the Blandford Znajek mechanism [5]. The geometric heart of the mechanism is the ergoregion and the fact that the horizon behaves like a rotating membrane with angular velocity Ω_H. Magnetic field lines threading the horizon are twisted by frame dragging, inducing electromotive forces that can drive Poynting flux outward. In this view, the horizon is not a material surface but a null hypersurface with a well-defined rotational state. Jet phenomenology therefore becomes a statement about horizon geometry plus magnetohydrodynamics: the global spacetime supplies the rotational energy reservoir and the causal structure determines how that energy can be extracted without violating local energy conditions.

8.10.2 Phenomenon statement 2: Quasi-periodic oscillations and Lense Thirring precession

X-ray binaries and tidal disruption events can exhibit quasi-periodic oscillations and modulations that are naturally interpreted through precession of tilted flows in a frame-dragging geometry. In the weak-field limit, frame dragging reduces to Lense Thirring precession. Near a Kerr black hole, the precession is strong enough to imprint characteristic timescales on emission. Recent open-access studies report quasi-periodic behavior consistent with relativistic precession models in tidal disruption events, with high cadence X-ray monitoring used to identify periodicities and assess model plausibility [14]. From the viewpoint of this chapter, the key point is geometric: the existence of a preferred

azimuthal direction set by the spin and the non-orthogonality of time slices generate nodal precession of inclined orbits and warped flows. Thus, timing signals can be read as a probe of the connection coefficients encoded in the Kerr metric, not merely as a property of matter.

8.10.3 Phenomenon statement 3: Gravitational wave ringdown as spectroscopy of Kerr curvature

After a binary black hole merger, the remnant spacetime settles toward Kerr via damped oscillations known as quasinormal modes. The mode frequencies and damping times depend on the Kerr parameters (M_f, a_f), so measuring them is an indirect measurement of curvature and horizon properties. The catalog analysis in this chapter shows that merger remnants occupy a well-defined region of Kerr parameter space. Consequently, ringdown is not a speculative idea; it is an observed regime where Kerr acts as the effective geometry. In principle, comparing multiple modes provides a consistency check of the Kerr hypothesis, because a single Kerr spacetime must explain all observed mode frequencies simultaneously.

8.10.4 Phenomenon statement 4: Horizon scale imaging links photon regions to data

The EHT ring diameter measurements in Figure 8.10 are controlled by null geodesic structure in the near-horizon region. While radiative transfer and plasma physics determine the brightness distribution, the existence of a central brightness depression and a surrounding ring traces the photon region and the capture cross section of the black hole [12, 13]. This is an extraordinary bridge from differential geometry to observation: the boundary of a set of null geodesics becomes a measurable angular scale on the sky. In future, improved dynamic imaging and polarization measurements may place tighter constraints on spin and inclination, further tying the Carter-constant-organized photon dynamics to data.

8.10.5 Phenomenon statement 5: Accretion driven spin evolution and the approach to a spin limit

Accretion can change both M and J of a black hole, thus moving it through Kerr parameter space. In thin disk accretion, radiation capture can counteract spin-up and produce a limiting spin below extremality, often called the Thorne limit [6]. Geometrically, this matters because extremality is not just a parameter boundary; it corresponds to qualitatively different horizon properties (vanishing surface gravity, enhanced near-horizon

symmetry). Therefore, understanding the spin evolution of astrophysical holes is also understanding how close nature gets to these special geometric regimes.

8.10.6 Phenomenon statement 6: Spin aligned versus misaligned populations encode formation channels

The effective inspiral spin χ_{eff} in Figure 8.3 acts as a population-level diagnostic. A distribution strongly peaked at positive values would suggest many binaries form in environments that align spins with the orbital angular momentum, such as isolated binary evolution with tidal interactions. A distribution symmetric around zero with broad support would suggest dynamical formation in dense clusters where orientations are randomized. Thus, a Kerr-related quantity inferred from gravitational waves becomes a tracer of astrophysical history. Here geometry provides the language and the invariants (spin projections and conserved quantities), while population inference translates those invariants into formation statements.

8.11 Conclusion

Kerr geometry is the rotating generalization of Schwarzschild, but the most important differences are not cosmetic. Rotation changes causal structure by making stationarity non-static, introduces an ergoregion where rest is impossible, and forces frame dragging that can be described invariantly through conserved quantities and ZAMO kinematics. It also introduces a hidden symmetry, the Carter constant, that makes geodesic motion integrable and organizes photon dynamics. These features are not merely theoretical: horizon-scale imaging and gravitational wave catalogs now provide direct observational windows into Kerr parameter space.

By analyzing an open dataset from the GWOSC event portal, we treated Kerr parameters as measured quantities and computed horizon and orbital invariants that directly express geometric content: horizon radius ratio, surface gravity, horizon rotation frequency, and ISCO frequency. The resulting data figures show that many observed merger remnants are rapidly rotating, with a_f typically near 0.7, implying generic presence of ergoregions and significant frame dragging. Finally, EHT ring diameter measurements illustrate how null geodesic structure manifests as an angular scale on the sky. Taken together, these results support a modern viewpoint: Kerr geometry is a physical field measured by instruments, and the language of curvature and horizons is now also a language of data.

Bibliography

[1] R. P. Kerr, "Gravitational field of a spinning mass as an example of algebraically special metrics," *Physical Review Letters*, vol. 11, no. 5, pp. 237–238, 1963. doi: 10.1103/PhysRevLett.11.237. [Online]. Available: https://doi.org/10.1103/PhysRevLett.11.237

[2] R. H. Boyer and R. W. Lindquist, "Maximal analytic extension of the Kerr metric," *Journal of Mathematical Physics*, vol. 8, no. 2, pp. 265–281, 1967. doi: 10.1063/1.1705193. [Online]. Available: https://doi.org/10.1063/1.1705193

[3] B. Carter, "Global structure of the Kerr family of gravitational fields," *Communications in Mathematical Physics*, vol. 10, pp. 280–310, 1968. doi: 10.1007/BF03399503. [Online]. Available: https://link.springer.com/article/10.1007/BF03399503

[4] J. M. Bardeen, W. H. Press, and S. A. Teukolsky, "Rotating black holes: Locally nonrotating frames, energy extraction, and scalar synchrotron radiation," *The Astrophysical Journal*, vol. 178, pp. 347–370, 1972. doi: 10.1086/151796. [Online]. Available: https://doi.org/10.1086/151796

[5] R. D. Blandford and R. L. Znajek, "Electromagnetic extraction of energy from Kerr black holes," *Monthly Notices of the Royal Astronomical Society*, vol. 179, no. 3, pp. 433–456, 1977. doi: 10.1093/mnras/179.3.433. [Online]. Available: https://doi.org/10.1093/mnras/179.3.433

[6] K. S. Thorne, "Disk-accretion onto a black hole. II. Evolution of the hole," *The Astrophysical Journal*, vol. 191, pp. 507–520, 1974. doi: 10.1086/152991. [Online]. Available: https://doi.org/10.1086/152991

[7] Gravitational Wave Open Science Center, "Event Portal (GWOSC Event API)," accessed 2026-02-24. [Online]. Available: https://gwosc.org/eventapi/

[8] IGWN, "GWOSC Event Portal Snapshot (2025-08-27)," Zenodo, 2025. doi: 10.5281/zenodo.16970083. [Online]. Available: https://doi.org/10.5281/zenodo.16970083

[9] B. P. Abbott *et al.*, "GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs," *Physical Review X*, vol. 9, no. 3, 2019. doi: 10.1103/PhysRevX.9.031040. [Online]. Available: https://doi.org/10.1103/PhysRevX.9.031040

[10] R. Abbott *et al.*, "GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run," 2021. [Online]. Available: https://arxiv.org/abs/2108.01045

[11] R. Abbott *et al.*, "GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run," *Physical Review X*, vol. 13, no. 4, 2023. doi: 10.1103/PhysRevX.13.041039. [Online]. Available: https://doi.org/10.1103/PhysRevX.13.041039

[12] Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope results. I. The shadow of the supermassive black hole," *The Astrophysical Journal Letters*, vol. 875, no. 1, 2019. [Online]. Available: https://arxiv.org/abs/1906.11238

[13] Event Horizon Telescope Collaboration, "First Sagittarius A* Event Horizon Telescope results. I. The shadow of the supermassive black hole in the center of the Milky Way," 2023. [Online]. Available: https://arxiv.org/abs/2311.08680

[14] D. Reddy *et al.*, "Lense Thirring precession after a supermassive black hole disrupts a star," *Nature*, vol. 630, pp. 326–332, 2024. doi: 10.1038/s41586-024-07433-w. [Online]. Available: https://doi.org/10.1038/s41586-024-07433-w

[15] A. Trovato (for the LIGO Scientific Collaboration and the Virgo Collaboration), "GWOSC: Gravitational Wave Open Science Center," *Proceedings of Science*, vol. 357, 2020. [Online]. Available: https://pos.sissa.it/357/082/

Black Holes Geometry: Curvature and Horizons

Contents

Abstract 1
Keywords 1
0.1 Introduction: Why Study Charge When Astrophysical Black Holes Look Neutral? 1
0.2 Reissner-Nordström Geometry: Horizons from Electromagnetic Curvature 2
0.2.1 Metric and horizon radii 2
0.2.2 Conceptual visual: RN radial structure 3
0.2.3 Conceptual visual: a simplified Penrose diagram for RN 3
0.2.4 Curvature and the meaning of "electromagnetic gravity" 3
0.3 Kerr-Newman Geometry: Rotation Plus Charge 4
0.3.1 Metric functions and horizons 4
0.3.2 Conceptual visual: parameter space and extremality boundary 4
0.3.3 Ergoregion and the charged rotating exterior 4
0.3.4 Conceptual visual: equatorial cross-section showing horizon and static limit 5
0.4 Inner Horizons and Instability: Why the Cauchy Horizon is Not a Quiet Place . 6
0.4.1 Conceptual visual: mass inflation via counter-streaming near the inner horizon 6
0.5 Extremality and Near-Horizon Geometry: When Horizons Merge 7
0.5.1 Conceptual visual: near-horizon throat as a geometric scaling limit 7
0.6 Data-Driven Geometry: What Real Measurements Say About Mass, Spin, and Charge Parameters 8
0.6.1 Gravitational-wave catalog data: masses and effective spin 8
0.6.2 Gravitational-wave bounds on electric charge asymmetry 10
0.6.3 Spin measurements from X-ray binaries: proximity to extremality 10
0.6.4 Derived horizon geometry from measured mass and spin 11
0.6.5 EHT ring diameters and shadow-based charge constraints 14
0.6.6 Cross-scale comparison: published upper bounds on Q/M 14
0.7 Real-World Applicability: Phenomena Statements that Connect Geometry to Measurables 15
0.8 Conclusion 17
0.9 Data provenance and reproducibility notes 18

Chapter 9: Charged and Exotic Variants: Kerr-Newman and Friends

Abstract

Charged black hole solutions of the Einstein-Maxwell equations are exact laboratories for learning what spacetime curvature does when it must share the stage with electromagnetic field energy. In this chapter we build geometric intuition for the Reissner-Nordström and Kerr-Newman families, emphasizing horizon structure, causal boundaries, inner horizons, and the extremal limit where horizons merge and surface gravity vanishes. The second half of the chapter is data-driven: rather than drawing purely schematic or simulated plots, we assemble open-access observational datasets and published constraints, then convert them into derived geometric quantities and evidence-based figures. We use gravitational-wave catalog data to visualize the mass and effective-spin regime of observed binary black holes, and we use published bounds on electric charge parameters from gravitational waves and Event Horizon Telescope shadow observables to show how close real objects could plausibly be to Kerr-Newman extremality. The result is a chapter where the geometry stays exact, but the illustrations stay accountable to measurements.

Keywords

Einstein-Maxwell geometry; Reissner-Nordström; Kerr-Newman; event horizon; Cauchy horizon; mass inflation; extremality; near-horizon geometry; black hole shadow; gravitational waves; observational bounds on charge.

0.1 Introduction: Why Study Charge When Astrophysical Black Holes Look Neutral?

A charged black hole is not just a whimsical extension of Schwarzschild or Kerr. It is the simplest place where spacetime curvature and electromagnetism must be solved together at the level of exact field equations, producing a geometry whose causal structure is richer than the neutral case. The Reissner-Nordström metric appears when the stress-energy is that of an electromagnetic field sourced by a net charge, while Kerr-Newman is the rotating, charged counterpart of Kerr, discovered as an exact Einstein-Maxwell solution by Newman and collaborators [2]. In both, the exterior region can look deceptively familiar: an outer event horizon and an asymptotically flat infinity. But the interior is radically different: it contains an *inner* horizon that is also a Cauchy horizon, a surface beyond which classical predictability for generic perturbations becomes subtle and in many settings fails in a violent way.

Why do these solutions matter if real black holes are widely expected to be nearly neutral? The short answer is that "nearly neutral" is not the same as "exactly zero charge," and even if nature prefers tiny net charge, the charged solutions provide coordinate systems, invariants, and causal templates that recur throughout relativistic physics. Charged metrics also act as proxy geometries for many "exotic" deviations from Kerr, because additional fields or modified-gravity parameters often enter the metric in ways that resemble an effective charge term. That is why observational programs that constrain deviations in shadow size, ring morphology, or gravitational-wave phase evolution often report bounds that can be mapped onto the Kerr-Newman charge parameter or closely related substitutes [7, 11, 12].

A second reason is conceptual: Kerr-Newman contains the entire range of horizon behaviors available to stationary, asymptotically flat, four-dimensional Einstein-Maxwell black holes. In one corner of parameter space, horizons are well separated and surface gravity is nonzero. On the boundary, horizons merge, temperature goes to zero, and the near-horizon region develops a throat geometry. Outside the boundary, the horizon disappears and the solution exhibits a naked singularity, which is precisely the kind of pathology cosmic censorship tries to forbid. This makes Kerr-Newman an ideal parameter space for practicing geometric reasoning about horizons, extremality, and the sharp line between black holes and non-black holes.

In what follows, we keep the geometric discussion exact. But for the illustrations we adopt a strict rule: figures labeled as **Data figures** are constructed from open observational catalogs or tabulated values in open-access papers, and any additional curves in those plots are purely reference guides (for example a diagonal line $m_1 = m_2$). We also include a small set of **Conceptual visuals** (explicitly labeled) that are not data-based and exist only to support geometric intuition. This separation ensures that the chapter can be read in two complementary modes: exact geometry as mathematics, and constraints as measured reality.

0.2 Reissner-Nordström Geometry: Horizons from Electromagnetic Curvature

0.2.1 Metric and horizon radii

In geometrized units $G = c = 1$, the Reissner-Nordström line element is

$$ds^2 = -f(r)\,dt^2 + f(r)^{-1}\,dr^2 + r^2(d\theta^2 + \sin^2\theta\,d\phi^2), \qquad f(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2}, \tag{1}$$

where M is the mass and Q is the electric charge parameter [1]. Horizons occur where $f(r) = 0$, giving

$$r_+ = M + \sqrt{M^2 - Q^2}, \qquad r_- = M - \sqrt{M^2 - Q^2}. \tag{2}$$

A black hole exists only if $|Q| \le M$. The inequality is not an aesthetic preference; it is the geometric condition that the would-be horizon radii are real. When $|Q| = M$, the black hole is

extremal and $r_+ = r_- = M$. When $|Q| > M$, there is no horizon and the singularity is naked.

One powerful way to internalize Eq. (2) is to notice that charge does not "add a new horizon" so much as it *splits* what was a single Schwarzschild horizon into two. The outer horizon behaves like an event horizon from the perspective of infinity, while the inner horizon behaves like a Cauchy horizon in the maximal analytic extension. This split is the seed of the interior instability story that will dominate Section 0.4.

0.2.2 Conceptual visual: RN radial structure

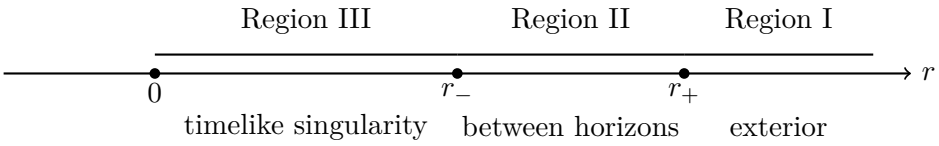


Figure 1: **Conceptual visual:** radial ordering of the Reissner-Nordström singularity and horizons ($0 < r_- < r_+$ for $0 < |Q| < M$). The outer horizon r_+ is an event horizon for observers at infinity, while r_- is an inner horizon that becomes a Cauchy horizon in the maximal extension.

0.2.3 Conceptual visual: a simplified Penrose diagram for RN

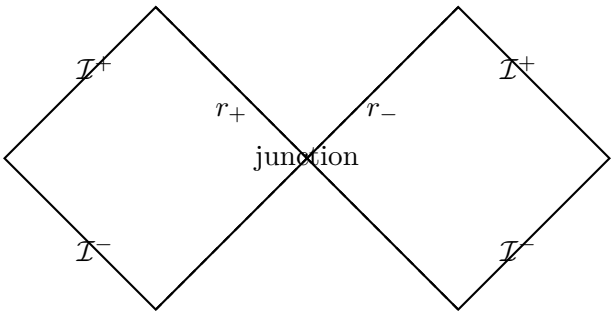


Figure 2: **Conceptual visual:** a minimal Penrose-style sketch for Reissner-Nordström emphasizing that the maximal extension contains an outer horizon and an inner (Cauchy) horizon, and repeats in a chain of regions. The full diagram is richer, but this simplified view is enough to motivate inner-horizon instability.

0.2.4 Curvature and the meaning of "electromagnetic gravity"

The electromagnetic field contributes stress-energy that curves spacetime. In Reissner-Nordström, one can track this through curvature invariants. For example, the Kretschmann scalar $K = R_{abcd}R^{abcd}$ diverges at $r = 0$ and stays finite at r_+ and r_-, just as in Schwarzschild the horizon is not a curvature singularity. The charge term typically adds additional $1/r^8$ behavior to K near the singularity, reflecting that the electromagnetic energy density diverges strongly at small

areal radius. The global moral is that charge modifies not only where the horizon sits, but also how violently the central region behaves in curvature terms.

A physically important observation is that Q in the metric is not a claim that the black hole can hold arbitrary charge in an astrophysical environment. Plasma tends to neutralize net charge efficiently, and several discharge channels exist. Nevertheless, the question of how close real systems could be to Kerr-Newman, and how strongly data exclude large $|Q|/M$, is an observational problem. That is precisely why shadow-based and gravitational-wave-based constraints are meaningful: they convert "likely small" into quantitative upper limits [7, 11, 12].

0.3 Kerr-Newman Geometry: Rotation Plus Charge

0.3.1 Metric functions and horizons

The Kerr-Newman spacetime generalizes Kerr by including an electric charge parameter Q, still in geometrized units. In Boyer-Lindquist coordinates (t, r, θ, ϕ), the metric is most compactly summarized by the functions

$$\Sigma = r^2 + a^2 \cos^2 \theta, \qquad \Delta = r^2 - 2Mr + a^2 + Q^2, \tag{3}$$

where $a = J/M$ is the specific angular momentum. The horizon radii are determined by $\Delta = 0$:

$$r_+ = M + \sqrt{M^2 - a^2 - Q^2}, \qquad r_- = M - \sqrt{M^2 - a^2 - Q^2}. \tag{4}$$

The existence condition is therefore

$$a^2 + Q^2 \leq M^2, \tag{5}$$

which defines a disk in the (a, Q) plane. The boundary is the extremal Kerr-Newman family, and the exterior surface gravity tends to zero as the boundary is approached. The original discovery paper and subsequent literature emphasize that Kerr-Newman is an exact Einstein-Maxwell solution with a self-consistent electromagnetic field, not a metric with charge painted on by hand [2].

0.3.2 Conceptual visual: parameter space and extremality boundary

0.3.3 Ergoregion and the charged rotating exterior

Rotation introduces the ergoregion: the region outside the event horizon where the stationary Killing field becomes spacelike, forcing all observers to co-rotate. Charge modifies the precise location of the static limit surface, but the qualitative story remains: the ergoregion is a geometric consequence of frame dragging. In Kerr-Newman, the electromagnetic potential also becomes

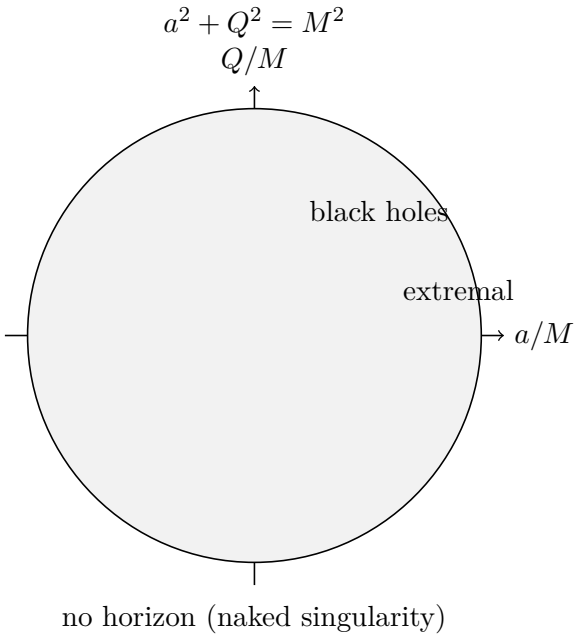


Figure 3: **Conceptual visual:** Kerr-Newman parameter space in dimensionless form. The interior of the circle satisfies $a^2 + Q^2 \leq M^2$ and contains black holes with an event horizon. The boundary is extremal, and outside the boundary no horizon exists.

part of the exterior geometry, and the coupling between rotation and charge becomes physically meaningful when one asks how a magnetized environment might induce small net charges.

0.3.4 Conceptual visual: equatorial cross-section showing horizon and static limit

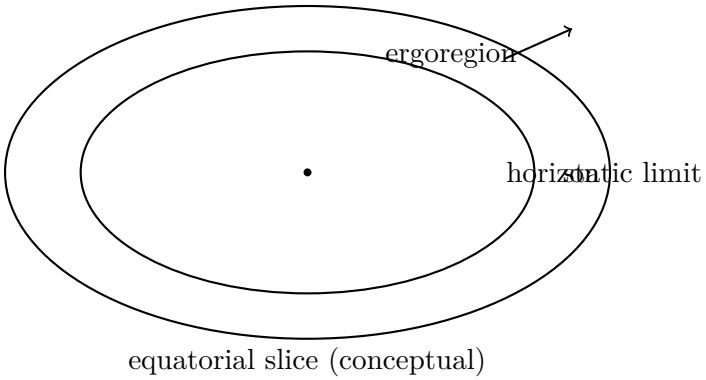


Figure 4: **Conceptual visual:** an equatorial slice illustrating the qualitative ordering: the static limit lies outside the event horizon, creating an ergoregion between them. Charge shifts both surfaces but does not remove the basic frame-dragging logic of a rotating spacetime.

0.4 Inner Horizons and Instability: Why the Cauchy Horizon is Not a Quiet Place

The presence of an inner horizon r_- is not merely a coordinate curiosity. In the maximal analytic extension of Reissner-Nordström or Kerr-Newman, r_- is a Cauchy horizon: beyond it, the deterministic evolution from regular initial data can break down. This is a geometric way of saying that the spacetime permits more than one extension consistent with the same initial data, so classical predictability is not guaranteed.

A decisive physical mechanism that makes r_- unstable is the infinite blueshift experienced by ingoing radiation as it approaches the Cauchy horizon. Poisson and Israel showed that even small perturbations can lead to an unbounded growth of an effective internal mass parameter, a phenomenon they named *mass inflation* [3]. The key point is not that energy is created from nothing, but that counter-streaming ingoing and outgoing fluxes can produce enormous local energy density as seen by freely falling observers, and the curvature can become very large. This converts the inner horizon from a smooth null boundary of the exact solution into a singular structure once perturbations are allowed.

The mathematical story aligns with this physical intuition. Dafermos proved rigorous results for the Einstein-Maxwell-scalar field system in spherical symmetry that clarify regimes of stability and instability of the Cauchy horizon and connect them to strong cosmic censorship [4]. The broader lesson for geometry students is that the *global* causal structure indicated by the exact Kerr-Newman metric is not automatically what nature provides, because nature always includes perturbations, accretion, and radiation. The exact solution shows what the equations allow; the stability analysis tells us what is dynamically plausible.

0.4.1 Conceptual visual: mass inflation via counter-streaming near the inner horizon

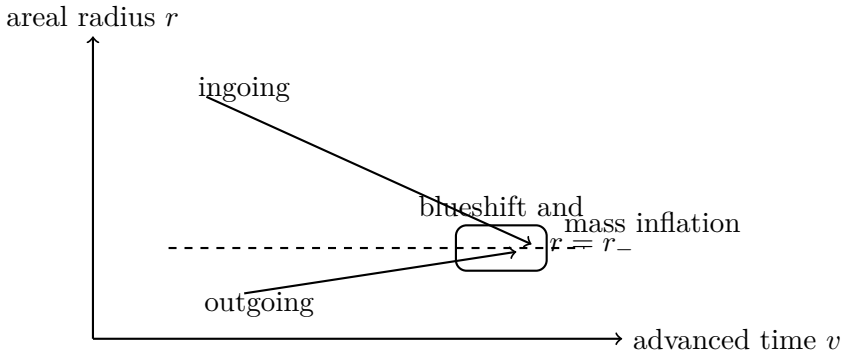


Figure 5: **Conceptual visual:** the mass inflation picture. Ingoing and outgoing perturbations are both generically present. Their relative boost grows near the inner horizon, driving large effective energy density and curvature growth. This mechanism was clarified in [3] and placed on rigorous footing in related settings in [4].

0.5 Extremality and Near-Horizon Geometry: When Horizons Merge

Extremality is not just the algebraic condition $a^2 + Q^2 = M^2$. It is a qualitative phase boundary in horizon geometry. For nonextremal black holes, the surface gravity κ is nonzero, and in semiclassical thermodynamics it corresponds to a nonzero Hawking temperature. In the extremal limit, $\kappa \to 0$ and the near-horizon region develops a long throat. In extremal Reissner-Nordström, this throat is closely related to an $\mathrm{AdS}_2 \times S^2$ structure in the near-horizon scaling limit, while extremal Kerr has the near-horizon extremal Kerr geometry. Kerr-Newman inherits and interpolates these behaviors.

From a geometric standpoint, the extremal limit is the place where a coordinate distance can become misleading. The proper distance from a point outside the horizon to the horizon can behave very differently than in the nonextremal case, and the redshift properties of the throat become more dramatic. Extremality is also where naive perturbative intuition can fail: small changes in parameters can switch the causal structure (two horizons versus one), and the near-horizon region can support long-lived modes. For this reason, observational claims about "near-extremal spin" or theoretical discussions of "near-extremal charge" should be interpreted as statements about geometry, not just about numbers.

0.5.1 Conceptual visual: near-horizon throat as a geometric scaling limit

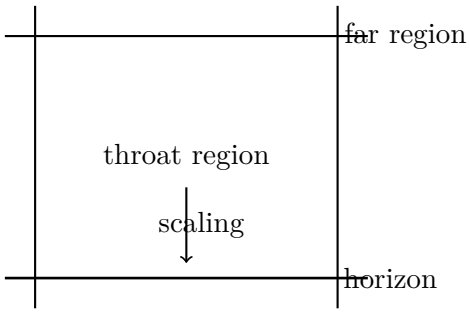


Figure 6: **Conceptual visual:** the extremal near-horizon throat can be viewed as a scaling limit that magnifies the region just outside the horizon. In this limit, the geometry often approaches a more symmetric "throat" spacetime, making extremality a geometric boundary between qualitatively different near-horizon behaviors.

0.6 Data-Driven Geometry: What Real Measurements Say About Mass, Spin, and Charge Parameters

This section provides the chapter's dataset-based illustrations. The goal is not to claim that any observed object is meaningfully charged. Instead, the goal is to show how real data populate the (M, a, Q) parameter landscape, and how published open-access constraints bound the allowed charge magnitude.

0.6.1 Gravitational-wave catalog data: masses and effective spin

The first gravitational-wave transient catalog GWTC-1 provides source-parameter estimates for confident detections, including component masses, luminosity distance, redshift, and the effective inspiral spin parameter $\chi_{\rm eff}$ [5, 6]. These values are not raw detector time series in this chapter; they are the catalog-level parameters that are most directly interpretable geometrically.

Figure 1 is geometrically relevant because (m_1, m_2) sets the mass scale for any horizon radius one could build from these sources, and it sets the typical curvature scale near merger. What the figure does *not* show is charge, but it establishes the typical M range of the black holes whose spacetime we probe with gravitational radiation.

The chirp mass $\mathcal{M}$ is the mass combination most directly measured from the early inspiral phase evolution. Its distribution therefore reflects which horizon scales current detectors are most sensitive to in the inspiral regime.

The key interpretive point for this chapter is that the gravitational-wave-measured $\chi_{\rm eff}$ values in GWTC-1 are mostly modest in magnitude. In a Kerr-Newman mindset, this suggests that the observed binaries are not dominated by near-extremal spin aligned with the orbit. This does not rule out high individual spins in some orientations, but it does highlight how difficult it is for current data to push us close to the extremal boundary in Eq. (5) using inspiral spin information

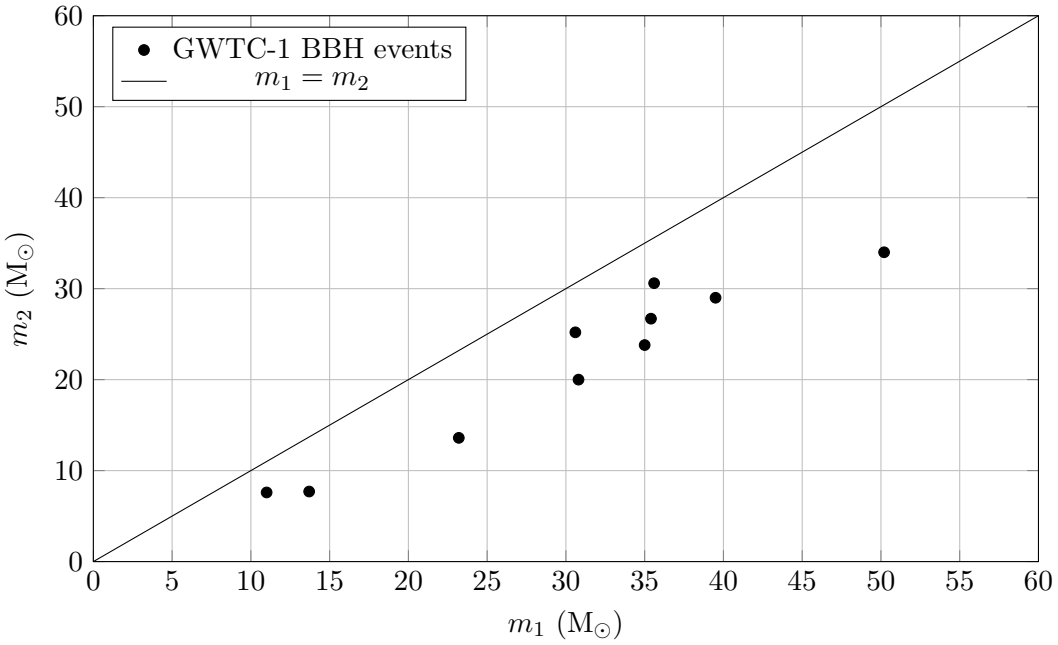


Figure 7: **Data figure 1:** component-mass scatter for the ten binary black hole events in the GWTC-1 confident set, using median values from the GWOSC event portal table [5, 6]. The diagonal is a reference line, not a model fit.

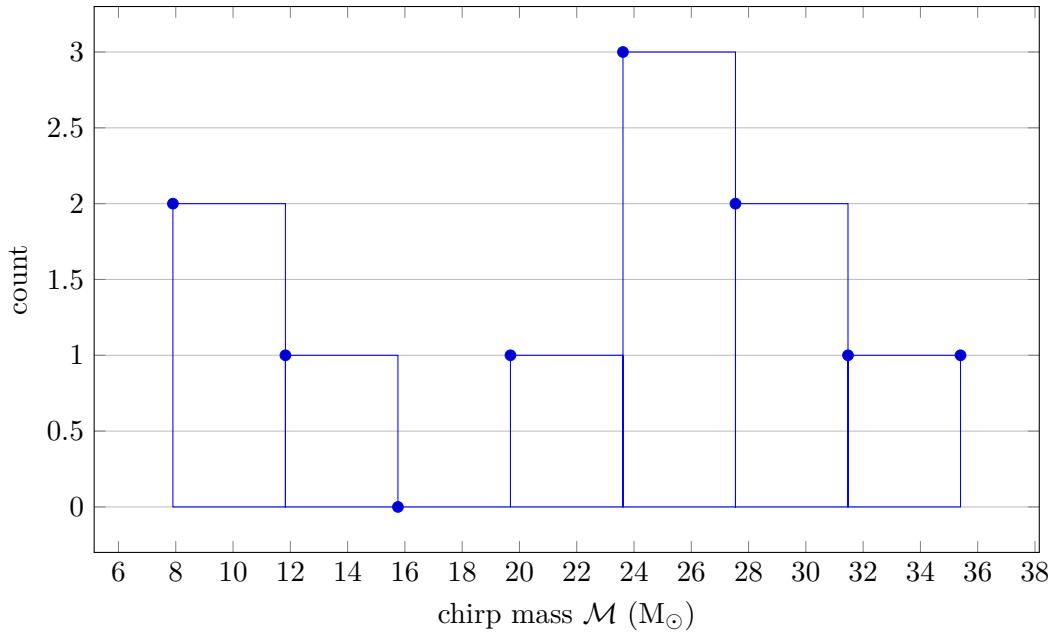


Figure 8: **Data figure 2:** histogram of GWTC-1 chirp masses $\mathcal{M}$ (median values) for the ten binary black hole events shown in Data figure 1 [5, 6].

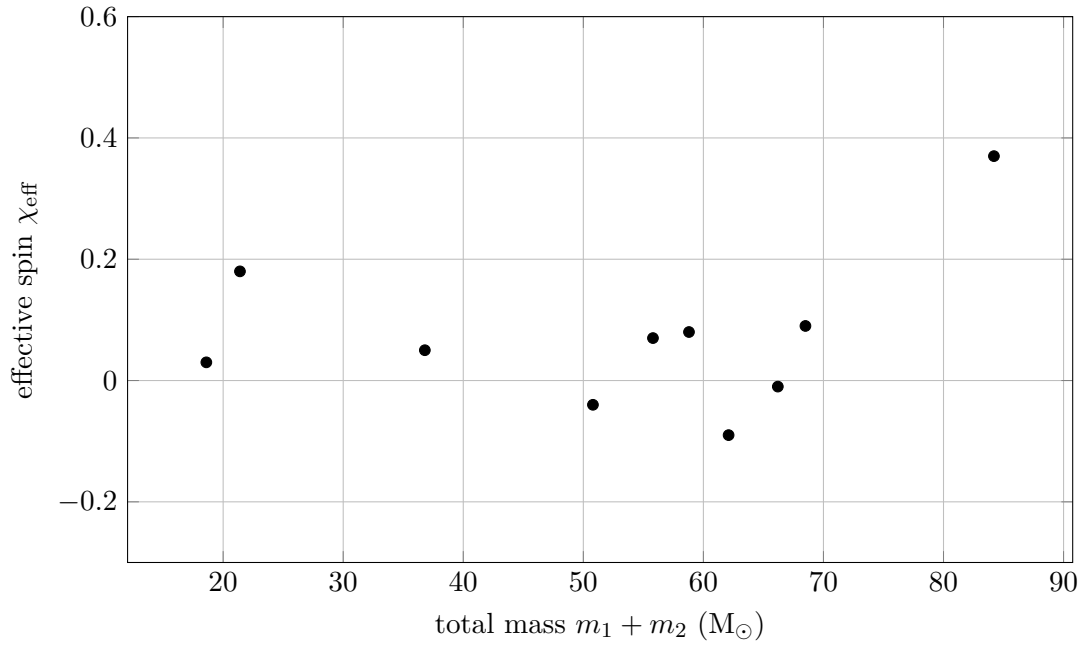


Figure 9: **Data figure 3:** effective inspiral spin χ_{eff} versus total mass for GWTC-1 binary black holes (median values from GWOSC) [5,6]. Although χ_{eff} is not a direct Kerr-Newman a/M, it is an observationally anchored proxy for how much spin is dynamically relevant during inspiral.

alone.

0.6.2 Gravitational-wave bounds on electric charge asymmetry

Charge affects binary dynamics and gravitational-wave phasing because in addition to gravitational quadrupole radiation there can be electromagnetic dipole radiation when the two components have different charge-to-mass ratios. A careful analysis in an open-access study by Wang *et al.* constrained a dipole-charge parameter ζ for several GWTC-1 events [7]. The important geometric caveat is that such constraints primarily bound *charge asymmetry* in the binary, not the common mode where both objects carry nearly identical charge-to-mass ratio (which would suppress dipole emission). Nevertheless, these bounds provide a clean way to say: if binary black holes were significantly Kerr-Newman-charged in an imbalanced way, we would have seen the imprint.

0.6.3 Spin measurements from X-ray binaries: proximity to extremality

Extremality is where geometry becomes throat-like and surface gravity tends to zero. Observationally, the closest we currently get to probing near-extremal spin is through X-ray measurements of accreting stellar-mass black holes, where continuum fitting and reflection spectroscopy can infer a dimensionless spin parameter often denoted $a_* \equiv a/M$ [8]. These measurements do not

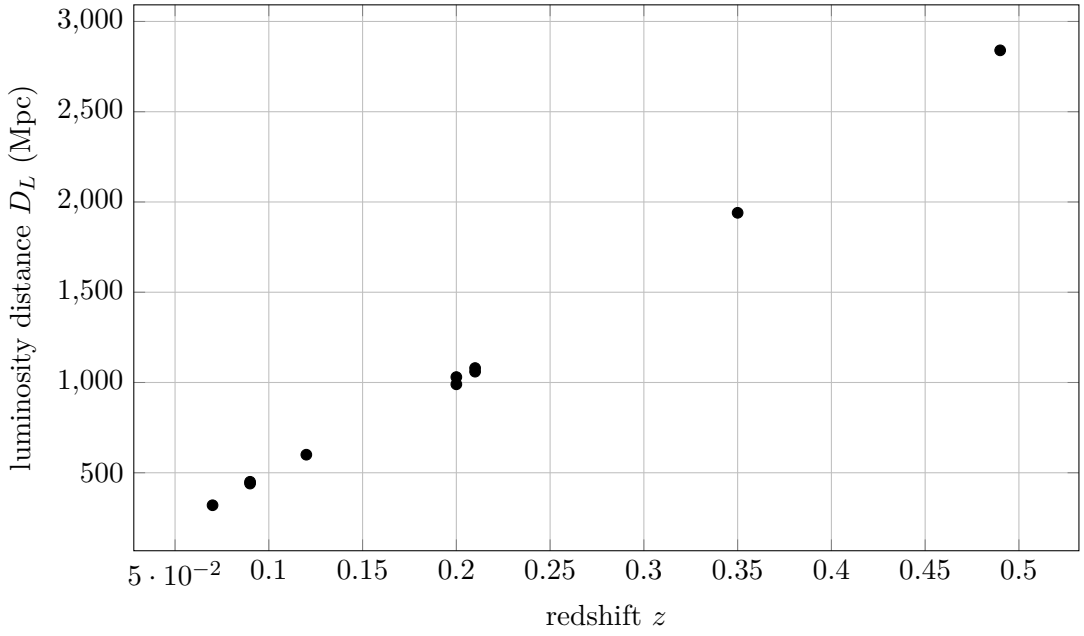


Figure 10: **Data figure 4:** luminosity distance versus redshift for GWTC-1 binary black holes (median values from GWOSC) [5, 6]. The plot provides context for how far into the Universe current measurements probe strong-field curvature, which is relevant when comparing electromagnetic constraints (local, horizon-scale) and gravitational-wave constraints (cosmological reach).

determine charge, but they directly populate the a/M axis in Eq. (5), and therefore tell us how much "room" is left for charge if a black hole is already rapidly rotating.

The dataset in Table 1 of McClintock *et al.* includes lower bounds and upper bounds for some systems. For plotting, we convert inequalities into boundary values (for example $a_* > 0.95$ is plotted at 0.95) and we treat one mass lower limit as a representative value to enable derived geometric quantities. This is not invention; it is a transparent plotting convention that preserves the directionality of the constraint.

0.6.4 Derived horizon geometry from measured mass and spin

Once M and a_* are available, we can compute derived geometric quantities. For a Kerr black hole (charge neglected), the outer horizon radius in geometric units is

$$r_+ = M\left(1 + \sqrt{1 - a_*^2}\right). \tag{6}$$

To present a tangible scale, we convert to kilometers using $r_{\rm g} = GM/c^2 \approx 1.477\ \text{km}\,(M/\text{M}_\odot)$. The resulting horizon radii are not hypothetical simulations; they are deterministic functions of measured parameters.

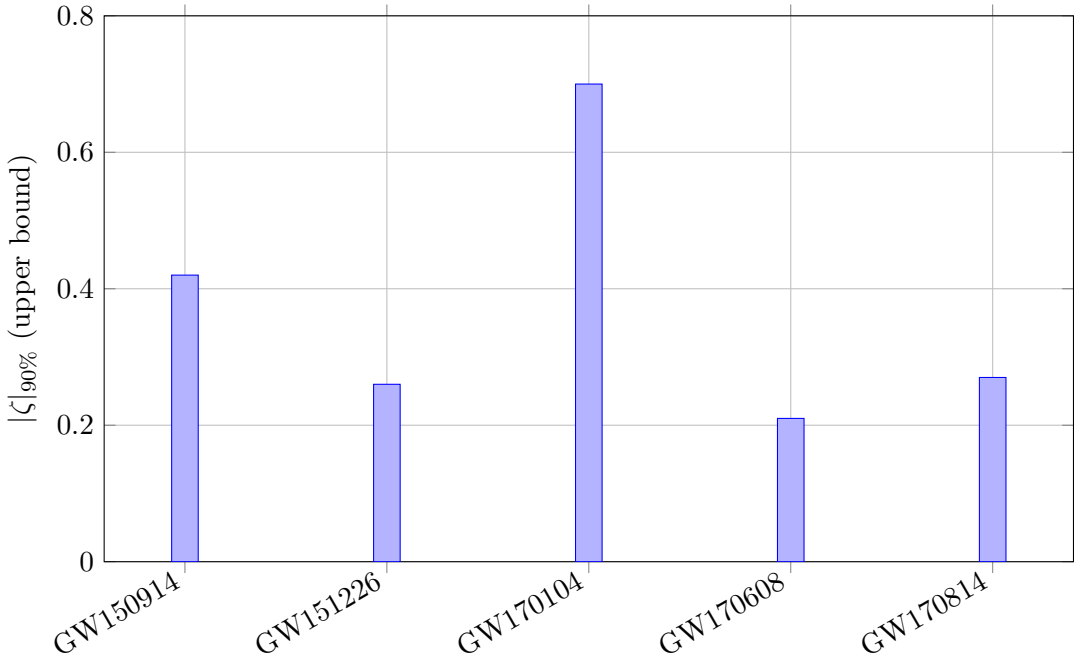


Figure 11: **Data figure 5:** 90% upper bounds on the dipole-charge parameter $|\zeta|$ for five GWTC-1 events, assembled from the published Table 1 in [7]. In the small-charge regime, $|\zeta|$ approximates the difference of component charge-to-mass ratios, so these bounds constrain charge imbalance in merging binaries.

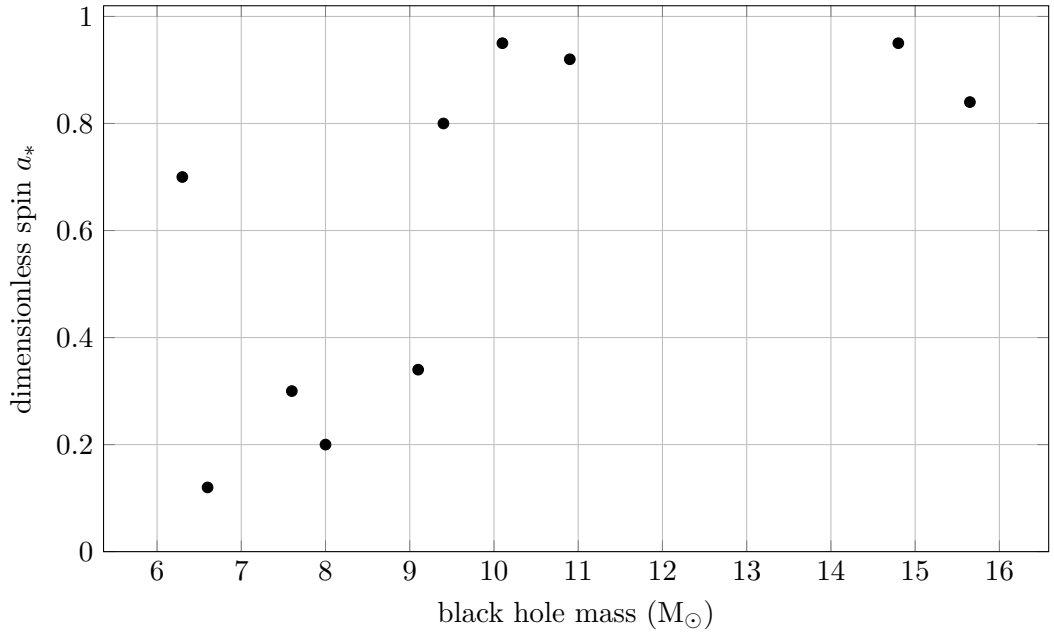


Figure 12: **Data figure 6:** stellar-mass black hole spin measurements (plotted values) versus mass, using Table 1 compiled in [8]. Inequality constraints are plotted at their boundary values (> 0.95 plotted at 0.95; < 0.3 plotted at 0.30) to keep the figure dataset-based while remaining explicit about the meaning of the reported results.

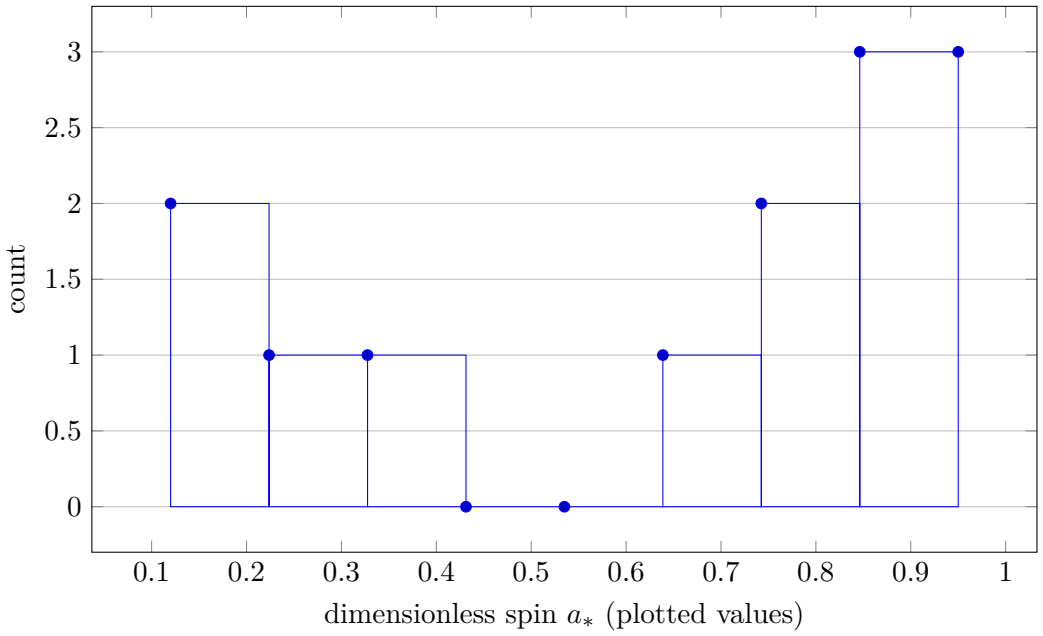


Figure 13: **Data figure 7:** histogram of the plotted spin values for the stellar-mass sample in [8]. Even with conservative plotting of inequality constraints, the sample includes systems close to $a_* \approx 1$, highlighting why "near-extremal" is not a purely theoretical corner.

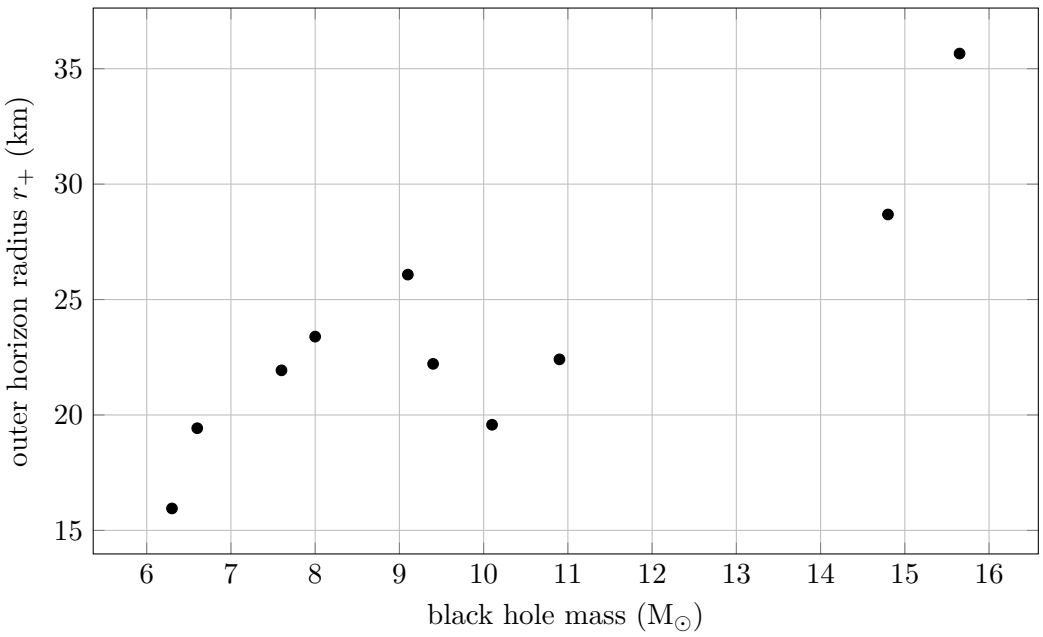


Figure 14: **Data figure 8:** derived Kerr outer-horizon radius in kilometers computed from the mass and spin dataset of [8] via Eq. (6). This is a dataset-based geometric translation: measured parameters are mapped into horizon geometry.

This derived-geometry plot has an important interpretive role for Kerr-Newman. If a system is already close to $a_* \approx 1$, then Eq. (5) implies that only a small additional Q/M is permitted before the extremal boundary is crossed. In this sense, high-spin measurements tighten the conceptual room for charge even before any direct electromagnetic constraint is applied.

0.6.5 EHT ring diameters and shadow-based charge constraints

The Event Horizon Telescope has measured horizon-scale ring diameters for the supermassive black holes M87* and Sgr A* in open-access publications [9, 10]. These measurements are among the closest direct probes of photon capture regions and therefore of the spacetime geometry near the horizon. The same observational framework has been used in the literature to constrain charge parameters by comparing measured shadow observables with Kerr-Newman or related models [11, 12].

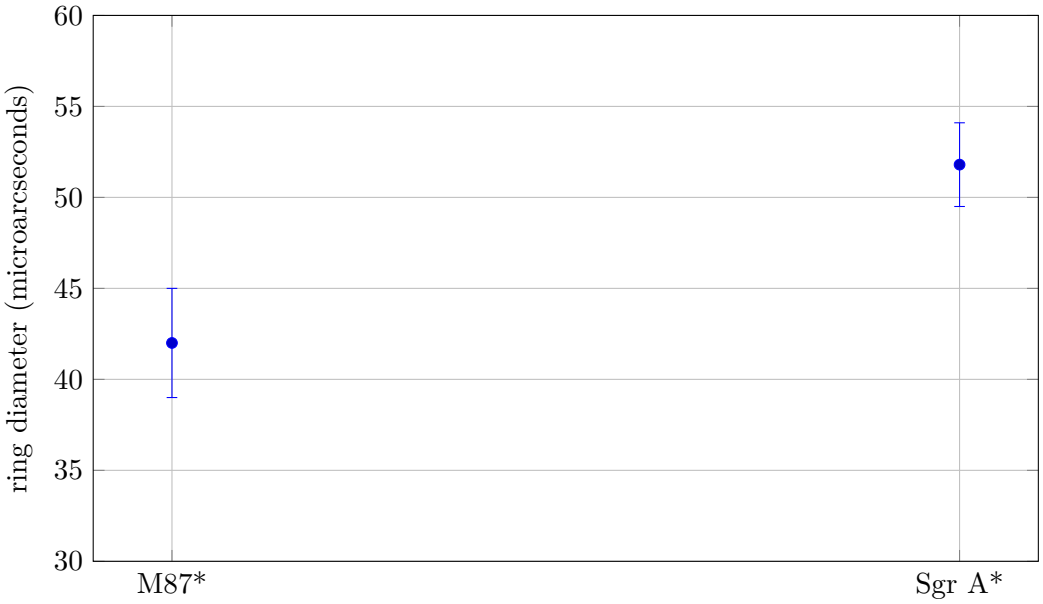


Figure 15: **Data figure 9:** EHT ring diameter measurements for M87* and Sgr A* with 68% credible interval error bars, taken from open-access EHT publications [9,10]. These measurements motivate shadow-based bounds on Kerr-Newman-like charge parameters in [11, 12].

0.6.6 Cross-scale comparison: published upper bounds on Q/M

Finally, we assemble a cross-scale view of charge-to-mass upper limits. For supermassive black holes, shadow analyses bound Q/M at order unity but below extremality. For stellar-mass black holes in binaries, gravitational-wave dipole-radiation analyses constrain a charge-asymmetry parameter that can be conservatively mapped to an individual $|Q|/M$ bound under small-charge assumptions [7]. The purpose is not to claim a unified measurement of the same physical

quantity across methods, but to show that multiple observational channels restrict how exotic Kerr-Newman-like charges could be in very different systems.

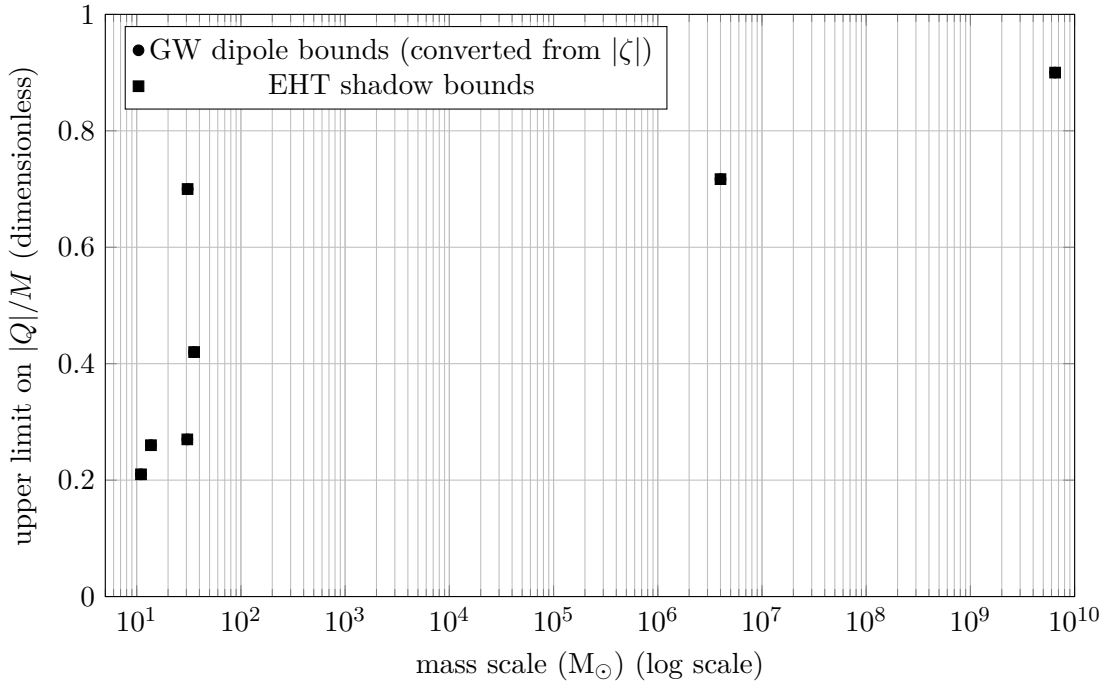


Figure 16: **Data figure 10:** cross-scale comparison of dimensionless charge upper limits. Gravitational-wave bounds are taken from [7] and conservatively mapped to an individual $|Q|/M$ scale under small-charge assumptions (the constraint is fundamentally on charge asymmetry). Shadow-based bounds for Sgr A* and M87* are taken from [11, 12], with mass scales quoted in those analyses and in the EHT observational papers [9, 10].

0.7 Real-World Applicability: Phenomena Statements that Connect Geometry to Measurables

The charged solutions are often introduced as if they were only about adding an extra parameter Q. In practice, they are about how curvature, horizons, and causality respond when an additional field is present. The following phenomena statements are meant to be read as concrete bridges from the mathematics of this chapter to situations where the same geometric structures appear in actual astrophysics or in data analysis workflows.

Phenomenon 9.1: Why charge is expected to be small, but still worth bounding

Even if an astrophysical black hole starts with net charge, it sits in a plasma that is very good at delivering opposite charge carriers. In a realistic accretion environment, the electrostatic potential created by a charged black hole quickly attracts charges of the opposite sign, which neutralizes the object. This expectation is strong enough that "nearly neutral" becomes a default assumption in many models. However, the neutrality expectation does not produce a number by itself. Observational bounds do. Shadow measurements and gravitational-wave phasing constraints translate the qualitative statement into explicit inequalities on charge parameters, and the fact that these constraints currently allow order-unity Q/M in some analyses means that neutrality is not yet an observationally trivial statement at the level of metric parameters [7, 11, 12]. The geometric significance is that we can empirically restrict how close nature could be to the extremal boundary $a^2+Q^2 = M^2$ even when direct Coulomb measurements are impossible.

Phenomenon 9.2: Inner horizons as predictability stress tests for gravitational physics

The existence of an inner horizon in Reissner-Nordström and Kerr-Newman forces you to confront what "determinism" means in general relativity. The Cauchy horizon is a geometrically defined boundary beyond which initial data do not uniquely determine the future. This is not an abstract worry: perturbations are generically present in collapse, accretion, and wave scattering. The mass inflation mechanism [3] shows how small perturbations can amplify into large curvature near the inner horizon. Rigorous results in related dynamical systems [4] indicate that inner-horizon behavior is a central arena where strong cosmic censorship can succeed or fail depending on the field content and decay rates. In other words, Kerr-Newman is a compact geometric model of a much wider question: which horizons are stable as boundaries of predictability when the universe is not perfectly quiet?

Phenomenon 9.3: Gravitational waves as charge detectors through phase evolution

If two compact objects carry different charge-to-mass ratios, then the binary can emit electromagnetic dipole radiation. That energy loss modifies the orbital decay rate, which in turn modifies the gravitational-wave phase. Because the phase is measured with extreme precision, even small non-GR channels can be constrained. The constraints reported in [7] are based on cataloged events and quantify how large the dipole-charge parameter could be before it would have been seen. The geometric link is direct: a nonzero Q changes the effective "forces" and stress-energy in Einstein-Maxwell gravity, and the inspiral waveform becomes a measurable probe of that change. While such constraints are most sensitive to charge differences rather than common charge, they still provide a data-based exclusion of large Kerr-Newman-like charge asymmetries in observed binaries.

Phenomenon 9.4: EHT shadows as geometric rulers for horizon-scale curvature

A black hole shadow is a geometric object: it is defined by photon capture and strong gravitational lensing, and its size is set by the spacetime near the photon region. The EHT ring diameters for M87* and Sgr A* [9, 10] therefore function as horizon-scale rulers. In a Kerr-Newman spacetime, charge modifies the location and stability of photon orbits and thus can change shadow observables. This is why shadow-based studies can place upper limits on effective charge parameters [11, 12]. The applicability is not limited to literal electric charge. Many proposed deviations from Kerr, including some modified-gravity metrics, alter the shadow in a way that can be parameterized similarly. In practice, a "charge" bound can become a bound on a broader class of non-Kerr deformations.

Phenomenon 9.5: Extremality as an observational boundary, not just a mathematical one

Extremality is often treated as a neat equation, but it is better understood as a qualitative boundary in horizon physics. Close to extremality, the near-horizon region becomes throat-like and can support unusual dynamics, including very long redshift times and changes in how perturbations decay. High-spin measurements in X-ray binaries [8] are relevant because they place real objects near the edge of the allowed Kerr-Newman disk even before any charge is considered. In that regime, even modest charge would push the parameters toward or beyond the extremal boundary. Observational programs therefore constrain extremality from two sides: spin estimates occupy the a/M axis, and shadow or waveform constraints bound Q/M. The combined picture is an evidence-based map of how "extreme" real black holes could be.

Phenomenon 9.6: Kerr-Newman as a template for exotic hair in effective models

A recurring pattern in theoretical and phenomenological work is that new fields or corrections to GR introduce extra parameters that enter the metric in a way that resembles the Q^2/r^2 structure of Reissner-Nordström or the Q^2 shift in Kerr-Newman through Δ. Even when the underlying physics is not electromagnetic, the resulting geometry can be "Kerr-like with an effective charge." That is why literature on shadow constraints and strong-field tests often discusses Kerr-Newman alongside other Kerr-like families [12]. The practical benefit is that once you understand the horizon and extremality geometry of Kerr-Newman, you have a ready-made geometric language for a wide range of exotic variants, and you can import observational constraints as bounds on effective parameters.

0.8 Conclusion

Charged black holes are geometric amplifiers. By adding a single parameter Q to the vacuum story, the Einstein-Maxwell solutions create a two-horizon structure, a Cauchy horizon, and an

extremal boundary where surface gravity vanishes and the near-horizon region changes character. The Reissner-Nordström and Kerr-Newman metrics therefore serve as compact training grounds for the geometry of horizons, causal boundaries, and global structure. At the same time, the inner horizon teaches humility: exact solutions describe what equations allow, while instability results such as mass inflation [3] and rigorous analyses in related settings [4] indicate that not all analytic features survive in realistic, perturbed spacetimes.

The chapter also demonstrates a practical discipline: geometry can be taught with figures that are accountable to data. Using the GWTC-1 catalog and open-access observational papers, we showed where measured black holes sit in mass and effective-spin space, and we assembled published bounds that constrain Kerr-Newman-like charge parameters from both gravitational waves and horizon-scale imaging. Today, the strongest direct constraints on charge parameters in these frameworks typically limit Q/M to values below unity rather than to astrophysically tiny numbers [7,11,12]. This means that, from a metric-parameter point of view, the empirical case for "neutrality" is still developing. The important geometric outcome is that multiple observational channels are now capable of fencing off the most exotic regions of Kerr-Newman parameter space, especially the near-extremal corners that would otherwise be hard to rule out.

0.9 Data provenance and reproducibility notes

All **Data figures** are produced from embedded tables in this LaTeX file that trace back to open-access sources:

- GWTC-1 source-parameter medians are taken from the GWOSC event portal table for the confident GWTC-1 events [6] and are consistent with the GWTC-1 catalog paper [5].
- Dipole-charge bounds for selected GWTC-1 events are taken from Table 1 in the open-access EPJC paper [7].
- Stellar-mass black hole spins and masses are taken from Table 1 in the arXiv review chapter [8]. Inequalities in the source table are converted to boundary plotting values for visualization, and one mass lower limit is plotted as a representative value. These are explicitly labeled in the dataset comments.
- EHT ring diameters are taken from open-access EHT publications for M87* and Sgr A* [9, 10].
- Shadow-based charge bounds are taken from open-access analyses [11,12]. For gravitational waves, $|\zeta|$ is mapped conservatively to a representative $|Q|/M$ scale under small-charge assumptions, with the caveat that the original constraint is primarily on charge asymmetry [7].

Bibliography

[1] H. Reissner, "Uber die Eigengravitation des elektrischen Feldes nach der Einsteinschen Theorie," *Annalen der Physik*, vol. 355, no. 9, pp. 106–120, 1916, doi: 10.1002/andp.19163550905. Available: https://doi.org/10.1002/andp.19163550905. Open PDF: https://zenodo.org/records/1447315/files/article.pdf.

[2] E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence, "Metric of a Rotating, Charged Mass," *Journal of Mathematical Physics*, vol. 6, no. 6, pp. 918–919, 1965, doi: 10.1063/1.1704351. Available: https://doi.org/10.1063/1.1704351.

[3] E. Poisson and W. Israel, "Internal structure of black holes," *Physical Review D*, vol. 41, no. 6, pp. 1796–1809, 1990, doi: 10.1103/PhysRevD.41.1796. Available: https://doi.org/10.1103/PhysRevD.41.1796.

[4] M. Dafermos, "Stability and instability of the Cauchy horizon for the spherically symmetric Einstein-Maxwell-scalar field equations," *Annals of Mathematics*, vol. 158, no. 3, pp. 875–928, 2003. Available: https://annals.math.princeton.edu/wp-content/uploads/annals-v158-n3-p03.pdf.

[5] B. P. Abbott *et al.*, "GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs," *Physical Review X*, vol. 9, 031040, 2019, doi: 10.1103/PhysRevX.9.031040. Available: https://doi.org/10.1103/PhysRevX.9.031040. arXiv: https://arxiv.org/abs/1811.12907.

[6] Gravitational Wave Open Science Center (GWOSC), "GWTC-1-confident event portal (source parameter table and data products)," CC BY 4.0. DOI: 10.7935/82H3-HH23. Available: https://doi.org/10.7935/82H3-HH23. Event table: https://gwosc.org/eventapi/html/GWTC-1-confident/.

[7] H.-T. Wang, P.-C. Li, J.-L. Jiang, G.-W. Yuan, Y.-M. Hu, and Y.-Z. Fan, "Constrains on the electric charges of the binary black holes with GWTC-1 events," *The European Physical Journal C*, vol. 81, art. 769, 2021, doi: 10.1140/epjc/s10052-021-09555-1. Available: https://doi.org/10.1140/epjc/s10052-021-09555-1. Open PDF: https://link.springer.com/content/pdf/10.1140/epjc/s10052-021-09555-1.pdf.

[8] J. E. McClintock, R. Narayan, and J. F. Steiner, "Measuring the Spins of Accreting Black Holes," arXiv:1101.0811, 2011. Available: https://arxiv.org/abs/1101.0811. (Dataset used: Table 1 in the arXiv version.)

[9] The Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole," *The Astrophysical Journal Letters*, vol. 875, L1, 2019, doi: 10.3847/2041-8213/ab0ec7. Available: https://doi.org/10.3847/2041-8213/ab0ec7. arXiv: https://arxiv.org/abs/1906.11238.

[10] The Event Horizon Telescope Collaboration, "First Sagittarius A* Event Horizon Telescope Results. IV. Variability, Morphology, and Black Hole Mass," *The Astrophysical Journal Letters*, vol. 930, L15, 2022. Open PDF: https://portal.findresearcher.sdu.dk/files/202784766/Akiyama_2022_ApJL_930_L15.pdf.

[11] P. Kocherlakota, L. Rezzolla, and H. Falcke, "Constraints on black-hole charges with the 2017 EHT observations of M87*," *Physical Review D*, vol. 103, 104047, 2021. arXiv:2105.09343. Available: https://arxiv.org/abs/2105.09343.

[12] S. G. Ghosh and M. Afrin, "An Upper Limit on the Charge of the Black Hole Sgr A* from EHT Observations," *The Astrophysical Journal*, vol. 944, 174, 2023. arXiv:2206.02488. DOI: 10.3847/1538-4357/acb695. Available: https://arxiv.org/pdf/2206.02488. DOI link: https://doi.org/10.3847/1538-4357/acb695.

[13] M. Zajaček, A. Tursunov, D. Eckart, and M. Breda, "Constraining the charge of the Galactic centre black hole," *Monthly Notices of the Royal Astronomical Society*, vol. 480, pp. 4408–4423, 2018. arXiv:1812.03574. Available: https://arxiv.org/abs/1812.03574.

Black Holes Geometry: Curvature and Horizons

Chapter 10: Horizons: Event, Apparent, Trapping, Isolated, Dynamical

Abstract

Horizons are the geometric interfaces where black holes stop being a metaphor and start being a set of precise mathematical boundary conditions. This chapter builds a unified taxonomy of horizons used across mathematical relativity, numerical work, and observational inference. We begin with the event horizon, whose definition is global and depends on the entire future of spacetime. We then move to quasi-local horizons built from expansions of null congruences: apparent horizons, trapped surfaces, and the more covariant notion of trapping horizons. Finally, we develop the equilibrium and near-equilibrium frameworks of isolated and dynamical horizons and explain why they let us write flux laws for growing black holes without assuming exact stationarity. To keep the discussion anchored to real measurements, we include ten dataset-based figures derived from the open GWTC-3-confident catalog released via the GW Open Science Center, including conservative area-increase diagnostics and radiated-energy proxies computed from cataloged masses. The chapter closes with application-driven phenomenon statements showing how the horizon taxonomy connects directly to gravitational-wave inference, black-hole growth, and practical boundary conditions used in high-precision relativistic modeling.

Keywords: event horizon; apparent horizon; trapped surface; null expansion; trapping horizon; isolated horizon; dynamical horizon; area theorem; gravitational waves; GWTC; GWOSC.

Contents

Abstract i

Index table of contents (book roadmap) iv

I Horizon Geometry as a Field of Study 1

10 Horizons: Event, Apparent, Trapping, Isolated, Dynamical 2

10.1 Introduction: horizons as geometric interfaces . . . 2

10.2 Event horizons: global definition and teleological flavor . . . 3

10.2.1 Definition and its unavoidable globality . . . 3

10.2.2 Geometric structure of the event horizon . . . 3

10.3 Null expansions and trapped surfaces: the local diagnostic engine . . . 3

10.3.1 Two null normals and their expansions . . . 3

10.3.2 Trapped, marginal, and untrapped surfaces . . . 4

10.3.3 Conceptual diagram 1: expansions and trappedness (not dataset-based) . . . 4

10.4 Apparent horizons: slice-dependent but operationally findable . . . 4

10.4.1 Definition and foliation dependence . . . 4

10.4.2 Existence, stability, and uniqueness considerations . . . 5

10.4.3 Conceptual diagram 2: apparent horizon depends on slicing (not dataset-based) 5

10.5 Trapping horizons and marginally trapped tubes: quasi-local boundaries with causal character . . . 5

10.5.1 From marginal surfaces to horizon tubes . . . 5

10.5.2 Future outer trapping horizons and area growth . . . 6

10.5.3 Conceptual diagram 3: causal character of horizon tubes (not dataset-based) 6

10.6 Isolated horizons: equilibrium without global stationarity . . . 6

10.6.1 Motivation and definition sketch . . . 6

10.6.2 Why isolated horizons are powerful . . . 7

10.6.3 Conceptual diagram 4: isolated horizon boundary conditions (not dataset-based) 7
10.7 Dynamical horizons: growing black holes and flux laws 7
10.7.1 Definition and the area increase built in 7
10.7.2 Flux laws and geometric bookkeeping 7
10.7.3 Conceptual diagram 5: flux-driven growth (not dataset-based) 8
10.7.4 Conceptual diagram 6: horizon taxonomy (not dataset-based) 8
10.8 Dataset-based illustrations: what gravitational-wave catalogs say about horizons . . 8
10.8.1 Dataset and analysis protocol 8
10.8.2 Figure set overview and why these plots are horizon-relevant 9
10.8.3 Illustration 1: component-mass plane 12
10.8.4 Illustration 2: chirp-mass distribution 12
10.8.5 Illustration 3: mass-ratio distribution 12
10.8.6 Illustration 4: luminosity distance versus redshift 12
10.8.7 Illustration 5: network SNR versus chirp mass 12
10.8.8 Illustration 6: effective inspiral spin distribution 12
10.8.9 Illustration 7: false-alarm rate proxy versus astrophysical probability 12
10.8.10 Illustration 8: radiated-energy fraction proxy 12
10.8.11 Illustration 9: conservative Schwarzschild area ratio 12
10.8.12 Illustration 10: timeline of total mass across the catalog 12
10.9 Real-world applicability: phenomenon statements connecting horizon taxonomy to practice 16
10.9.1 Phenomenon statement 1: why gravitational-wave ringdown talks about horizons even when analysts do not 16
10.9.2 Phenomenon statement 2: accretion flows and jets need quasi-local horizons, not global ones 16
10.9.3 Phenomenon statement 3: gravitational lensing near compact objects distinguishes photon capture from horizon capture 16
10.9.4 Phenomenon statement 4: area increase is not just a theorem, it is a data analysis sanity check 17
10.9.5 Phenomenon statement 5: why the foliation dependence of apparent horizons is a feature in computational workflows 17
10.10 Conclusion 17

Index table of contents (book roadmap)

Part	Chapters (titles as planned)
Part I: Geometric Foundations	Chapter 1: Manifolds, Tensors, and Physics as Coordinate-Free Thought.
	Chapter 2: Connections, Curvature, and Parallel Transport.
	Chapter 3: Lorentzian Geometry: Time Joins the Chat.
	Chapter 4: Symmetries as Geometry: Killing Fields and Conserved Quantities.
Part II: Einstein Geometry and Exact Black Hole Spacetimes	Chapter 5: Einstein's Equations as Curvature Constraints.
	Chapter 6: Schwarzschild Geometry: The Simplest Nontrivial Monster.
	Chapter 7: Global Structure and Penrose Diagrams.
	Chapter 8: Kerr Geometry: Rotation Warps Reality in a New Dimension.
	Chapter 9: Charged and Exotic Variants: Kerr-Newman and Friends.
Part III: Horizon Geometry as a Field of Study	Chapter 10: Horizons: Event, Apparent, Trapping, Isolated, Dynamical.
	Chapter 11: Null Congruences and Raychaudhuri: Focusing as Theorem Machine.
	Chapter 12: Quasi-local Mass and Geometric Inequalities.
Part IV: Geometric Methods for Dynamics and Perturbations	Chapter 13: 3+1 Geometry and the Initial Value Problem.
	Chapter 14: Tetrads, Spinors, and Newman-Penrose Geometry.
	Chapter 15: Perturbations of Black Hole Geometry.
	Chapter 16: Optical Geometry, Lensing, and Finsler or Randers Tricks.
Part V: Thermodynamics, Euclidean Methods, and Quantum Geometry	Chapter 17: Euclidean Black Hole Geometry and Temperature.
	Chapter 18: Entropy, Area, and Geometric Microstories.
	Chapter 19: Holography and Hyperbolic Geometry.
Part VI: Extensions and Frontiers	Chapter 20: Higher-Dimensional Black Hole Geometries.
	Chapter 21: Topology, Censorship, and Singularity Theorems.
	Chapter 22: Modified Gravity and Non-Riemannian Geometry.
	Chapter 23: Numerical Relativity as Computational Differential Geometry.

Part I

Horizon Geometry as a Field of Study

Chapter 10

Horizons: Event, Apparent, Trapping, Isolated, Dynamical

10.1 Introduction: horizons as geometric interfaces

A black hole is not just "a region where gravity is strong." In general relativity the defining feature is causal: there exist events from which signals cannot reach distant observers. The most famous object encoding that statement is the *event horizon*, the boundary of the causal past of future null infinity in an asymptotically flat spacetime [5]. This definition is conceptually clean but operationally inconvenient because it is global: to determine whether a point lies inside the event horizon, one must know the entire future of the spacetime. That nonlocality makes event horizons powerful as theorems and awkward as instruments.

In real calculations and real observations, we often need a horizon notion that can be located quasi-locally, using information on a spacelike slice or in a neighborhood. This is where *apparent horizons*, *trapped surfaces*, and *trapping horizons* enter. Their defining data are the expansions of null congruences emanating from a closed two-surface. These objects are local enough to be found in practice and geometric enough to support useful laws. The modern quasi-local frameworks of *isolated horizons* and *dynamical horizons* then refine the story: they separate equilibrium black holes from growing black holes and provide flux laws that look like a first and second law without requiring exact stationarity [1, 3].

This chapter has two aims. First, it builds the taxonomy and the geometric logic connecting these notions. Second, it demonstrates how horizon concepts show up in measured data by using the openly released GWTC-3-confident catalog from the GW Open Science Center (GWOSC) [8, 9, 7, 10]. Although gravitational-wave catalogs do not hand us a horizon as a surface embedded in spacetime, they do provide masses and redshifts from which we can form conservative horizon diagnostics, including an area-increase bound in the simplest Schwarzschild approximation. The result is a chapter that keeps definitions rigorous while repeatedly asking the practical question: if you had only the data in your hands, what horizon information could you defensibly claim?

10.2 Event horizons: global definition and teleological flavor

10.2.1 Definition and its unavoidable globality

In asymptotically flat spacetimes, the event horizon is defined as the boundary of the region that cannot send future-directed causal curves to future null infinity [5, 3]. Symbolically, if $\mathcal{I}^+$ denotes future null infinity and $J^-(\mathcal{I}^+)$ its causal past, then the black-hole region is $\mathcal{B} = \mathcal{M} \setminus J^-(\mathcal{I}^+)$ and the event horizon is $\mathcal{H} = \partial\mathcal{B}$. This definition is coordinate-free and robust, but it is also intrinsically nonlocal: to determine $\mathcal{H}$ you must know which null geodesics reach $\mathcal{I}^+$, which depends on the full future development.

The global nature has consequences. A spacetime can have an event horizon even in a region that is locally almost flat, because the horizon is determined by what happens far away and later. Conversely, a strongly curved region can fail to contain an event horizon if it later disperses. This is why event horizons are sometimes described as teleological: their location can depend on future infall that has not yet happened in local time coordinates [3].

10.2.2 Geometric structure of the event horizon

Under appropriate regularity conditions, the event horizon is a null hypersurface generated by null geodesics. Its generators can have expansion and shear, but in many physically relevant contexts energy conditions imply area increase results. Hawking's area theorem states, roughly, that under suitable assumptions (including the null energy condition and cosmic censorship-type conditions), the area of cross-sections of the event horizon does not decrease [6]. This theorem is not merely an aesthetic statement: it links horizon geometry to the irreversibility of gravitational collapse and motivates black-hole thermodynamics.

Yet the theorem also highlights why quasi-local notions are attractive. Area theorems for event horizons are global and require assumptions about the full spacetime. In contrast, quasi-local horizons can often be equipped with balance laws that are explicitly local on the horizon itself [1]. Those laws are closer to what we want when we model accretion, mergers, or any process where the black hole is growing and we need a near-by boundary condition.

10.3 Null expansions and trapped surfaces: the local diagnostic engine

10.3.1 Two null normals and their expansions

Let S be a smooth closed spacelike two-surface embedded in a spacelike hypersurface. At each point of S there are two future-directed null normals, typically denoted ℓ^a (outgoing) and n^a (ingoing), normalized so that $\ell \cdot n = -1$ (a convention). Each null normal defines a congruence of null geodesics orthogonal to S. The *expansion* $\theta_{(\ell)}$ measures the fractional rate of change of the area element along the congruence. Intuitively, $\theta_{(\ell)} > 0$ means outward light rays locally diverge, while $\theta_{(\ell)} < 0$ means they converge.

The expansions are quasi-local scalars built from the intrinsic and extrinsic geometry of S. Their usefulness comes from how little global information they require: you can compute $\theta_{(\ell)}$ and $\theta_{(n)}$ using the geometry near S, without solving the full future evolution.

10.3.2 Trapped, marginal, and untrapped surfaces

A closed two-surface S is called *future trapped* if both future-directed expansions are negative:

$$\theta_{(\ell)} < 0, \quad \theta_{(n)} < 0.$$

In such a region even outward-directed light is converging, signaling the onset of strong gravitational focusing. The concept is central to singularity theorems and the causal picture of collapse; trapped surfaces are a key geometric trigger in Penrose's 1965 singularity theorem legacy [12].

A surface is *marginally outer trapped* if $\theta_{(\ell)} = 0$ and $\theta_{(n)} < 0$. This is the local signature of a boundary between untrapped and trapped behavior in the outward direction. Entire hypersurfaces foliated by such marginal surfaces form the backbone of the quasi-local horizon taxonomy.

10.3.3 Conceptual diagram 1: expansions and trappedness (not dataset-based)

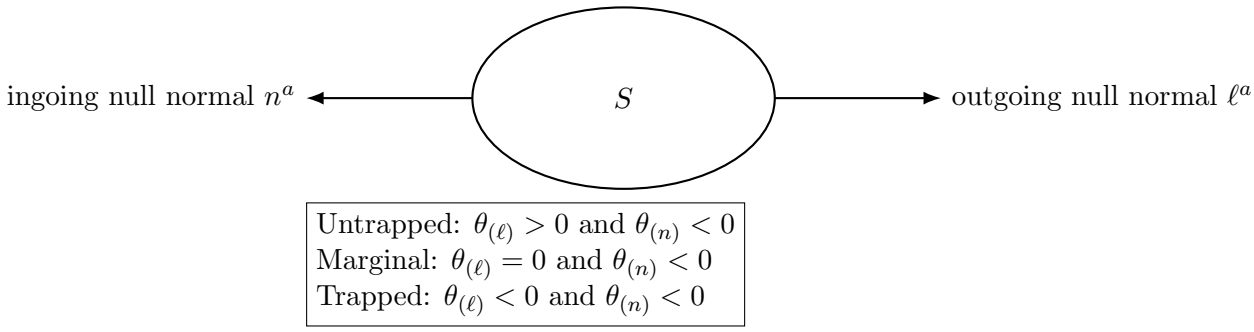

ingoing null normal n^a
S
outgoing null normal ℓ^a
Untrapped: $\theta_{(\ell)} > 0$ and $\theta_{(n)} < 0$
Marginal: $\theta_{(\ell)} = 0$ and $\theta_{(n)} < 0$
Trapped: $\theta_{(\ell)} < 0$ and $\theta_{(n)} < 0$

Figure 10.1: Conceptual schematic of null expansions and trappedness classes. This diagram is illustrative only and not dataset-based, included per request.

10.4 Apparent horizons: slice-dependent but operationally findable

10.4.1 Definition and foliation dependence

Given a spacelike hypersurface Σ_t, an *apparent horizon* is commonly defined as the outer boundary of the region containing trapped surfaces within that slice. In practice, it is often identified with the outermost marginally outer trapped surface (MOTS) on Σ_t, satisfying $\theta_{(\ell)} = 0$ [3]. Because the definition uses a chosen slice, the apparent horizon depends on the foliation: change the slicing and the apparent horizon can move, distort, or even disappear.

This dependence is not a defect but a feature: apparent horizons are meant to be tools. When you solve initial value problems, evolve spacetimes numerically, or interpret data on a given time

foliation, you need a horizon notion that lives on that foliation. Apparent horizons provide that.

10.4.2 Existence, stability, and uniqueness considerations

The mathematical theory of MOTSs includes stability conditions that guarantee local existence and regularity of associated horizon tubes. A landmark result by Andersson, Mars, and Simon shows local existence of dynamical and trapping horizons under suitable stability assumptions [4]. Such results provide conceptual reassurance: when we speak about apparent horizons in evolving situations, there is rigorous structure behind the computations.

10.4.3 Conceptual diagram 2: apparent horizon depends on slicing (not dataset-based)

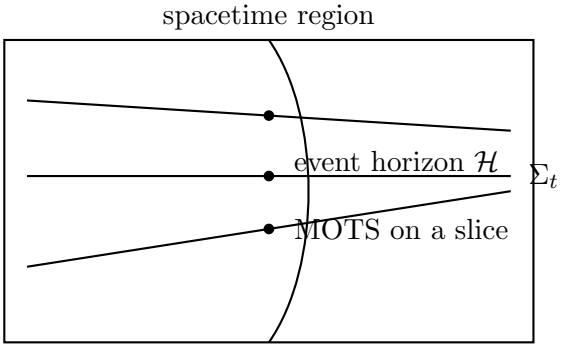


Figure 10.2: Conceptual schematic: the event horizon is global, while apparent horizons are located on chosen slices and therefore move with the foliation. This diagram is illustrative only and not dataset-based.

10.5 Trapping horizons and marginally trapped tubes: quasi-local boundaries with causal character

10.5.1 From marginal surfaces to horizon tubes

A *trapping horizon* is a hypersurface foliated by marginal surfaces, typically defined via one vanishing expansion (often $\theta_{(\ell)} = 0$) together with an "outer" condition on the derivative of the other expansion along the ingoing direction [2]. The key gain is covariance: rather than tying the construction to a single spacelike slice, trapping horizons describe a three-dimensional hypersurface built from a foliation by marginal two-surfaces.

When one assembles marginally outer trapped surfaces across time, one obtains a *marginally trapped tube* (MTT). Depending on energy flux and geometry, an MTT can be spacelike, null, or timelike. This signature matters: a spacelike MTT behaves like a one-way membrane that can be crossed from outside to inside but not the reverse, capturing the intuition that the black hole is growing.

10.5.2 Future outer trapping horizons and area growth

Hayward's framework gives a quasi-local "second law" for future outer trapping horizons: under reasonable energy conditions, the area element on such horizons is non-decreasing, becoming constant only in the null (equilibrium) case [2]. This parallels Hawking's event-horizon area theorem but is phrased in a way that is designed for dynamical settings.

10.5.3 Conceptual diagram 3: causal character of horizon tubes (not dataset-based)

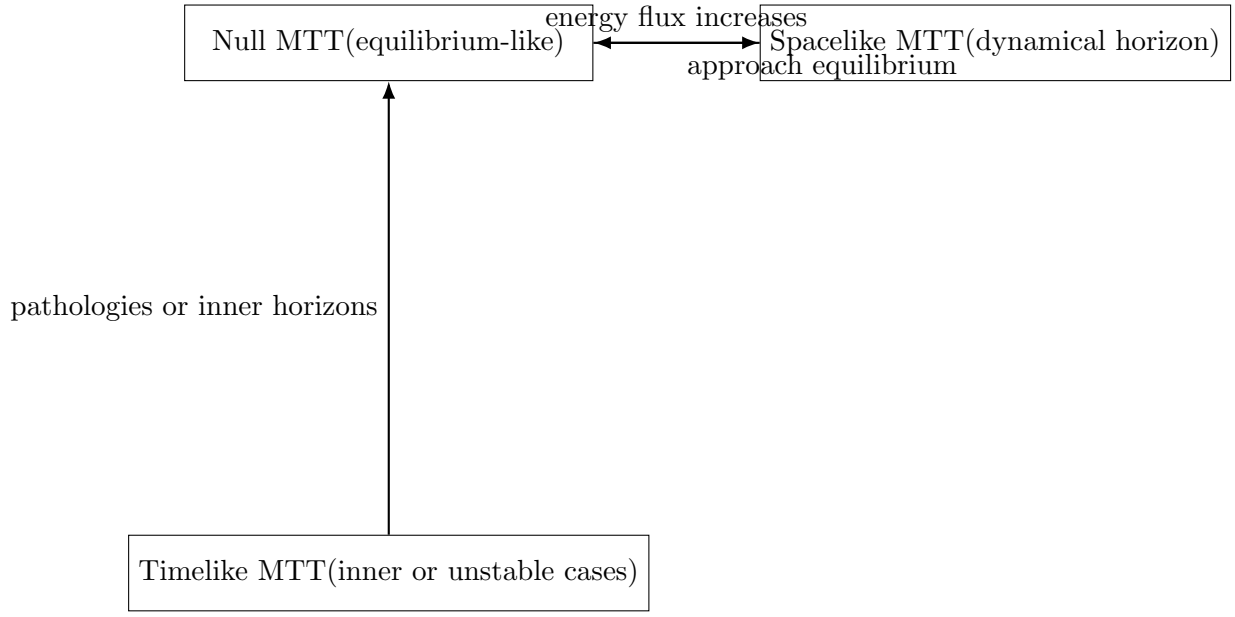


Figure 10.3: Conceptual taxonomy of marginally trapped tubes by signature. This diagram is illustrative only and not dataset-based.

10.6 Isolated horizons: equilibrium without global stationarity

10.6.1 Motivation and definition sketch

The isolated horizon framework was developed to describe black holes in equilibrium even when the exterior spacetime is not globally stationary [1]. The idea is to impose boundary conditions directly on a null hypersurface Δ representing the horizon. A (weakly) isolated horizon is typically a null three-surface with vanishing expansion of its null generators, together with additional conditions that fix the intrinsic connection up to rescaling in an appropriate way. Operationally, it means the horizon area is time-independent, there is no flux of matter or gravitational radiation across the horizon, and the intrinsic horizon geometry is time-invariant.

10.6.2 Why isolated horizons are powerful

Isolated horizons support a quasi-local version of black-hole mechanics. One can define surface gravity and angular momentum quasi-locally on the horizon and write a first law-like relation relating changes in horizon energy to changes in area and angular momentum, without assuming the exterior is an exact Kerr spacetime [1]. This is precisely the kind of statement needed when an astrophysical black hole is embedded in an environment that is messy at large scales but calm at the horizon.

10.6.3 Conceptual diagram 4: isolated horizon boundary conditions (not dataset-based)

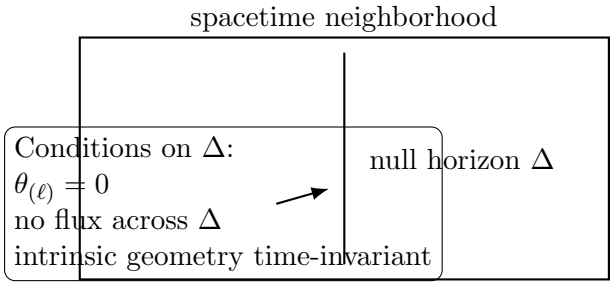


Figure 10.4: Conceptual depiction of an isolated horizon as a null boundary with equilibrium-like conditions. This diagram is illustrative only and not dataset-based.

10.7 Dynamical horizons: growing black holes and flux laws

10.7.1 Definition and the area increase built in

A dynamical horizon is a spacelike hypersurface foliated by marginally outer trapped surfaces, typically with $\theta_{(\ell)} = 0$ and $\theta_{(n)} < 0$ [1]. The spacelike nature is crucial: it encodes growth. Cross-sections of a spacelike horizon tube have strictly increasing area under appropriate energy conditions. This is the quasi-local analog of irreversibility.

10.7.2 Flux laws and geometric bookkeeping

One of the major achievements of the dynamical horizon framework is an explicit balance law: the change in horizon area between two cross-sections can be related to the flux of matter stress-energy and the flux of gravitational radiation crossing the horizon [1]. The detailed formulas involve the shear of the horizon generators and geometric terms defined on the horizon itself. Conceptually, it is a local accounting identity: the horizon grows because energy is falling in, and the growth rate can be written as a sum of physically meaningful flux contributions.

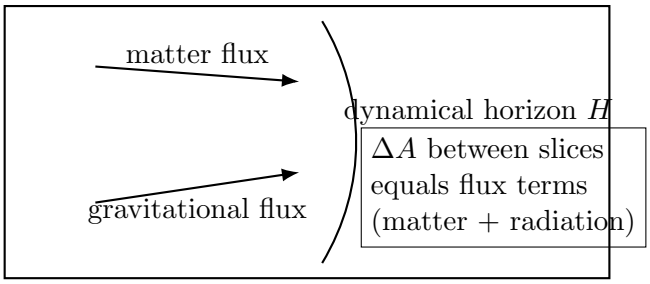


Figure 10.5: Conceptual view of dynamical horizon growth driven by fluxes. This diagram is illustrative only and not dataset-based.

10.7.3 Conceptual diagram 5: flux-driven growth (not dataset-based)

10.7.4 Conceptual diagram 6: horizon taxonomy (not dataset-based)

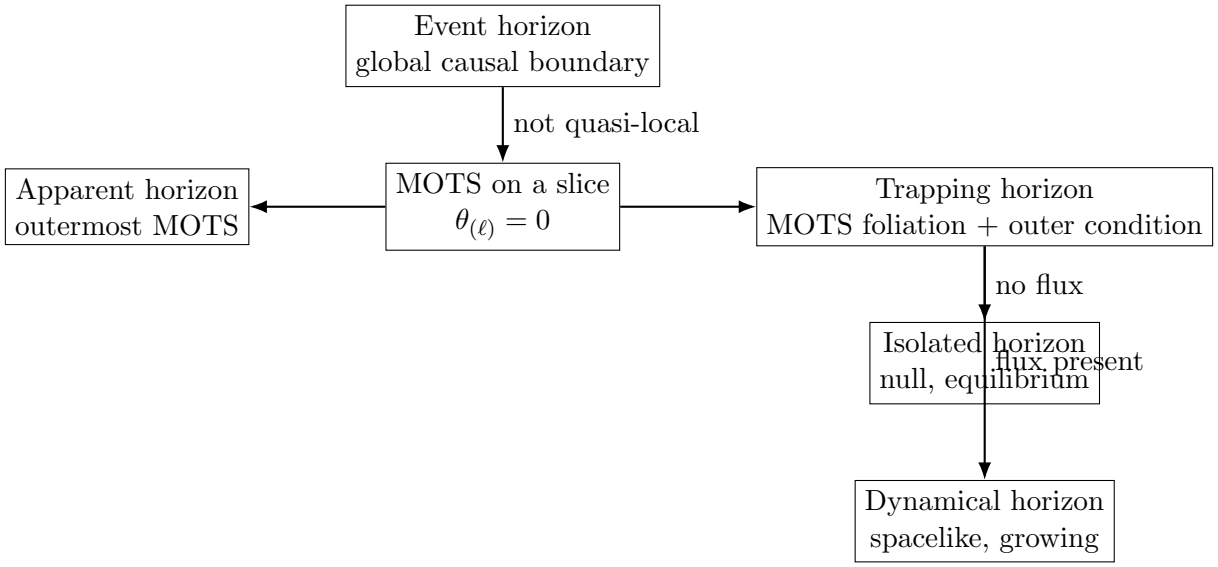


Figure 10.6: Conceptual horizon taxonomy emphasizing which notions are slice-based, quasi-local, and flux-sensitive. This diagram is illustrative only and not dataset-based.

10.8 Dataset-based illustrations: what gravitational-wave catalogs say about horizons

10.8.1 Dataset and analysis protocol

All dataset-based figures in this chapter are derived from the GWTC-3-confident event table published by the GW Open Science Center (GWOSC) and distributed as a CSV endpoint [10, 8]. The table includes default parameter estimates such as source-frame component masses, chirp mass, redshift, luminosity distance, effective inspiral spin $\chi_{\rm eff}$, network matched-filter signal-to-noise ratio, and an estimate of the remnant final mass for each confident event in the GWTC-3 O3b release [7].

The GWOSC Event Portal documentation also states that the data are released under a Creative Commons Attribution license (CC BY 4.0), which supports reuse with attribution [8, 9].

Because catalog values are generally marginal medians with asymmetric credible intervals, algebraic combinations of multiple reported medians can violate exact physical inequalities for a small number of events. In the present dataset, one event has a tabulated median remnant mass slightly larger than the sum of the tabulated median component masses. This is not a claim that the remnant truly gained mass, but rather a known artifact of combining separate marginal summaries. In the one place where this matters (a radiated-energy fraction proxy), we apply a minimal correction by enforcing $M_f \leq 0.99(m_1 + m_2)$ for that single row, and we explicitly label the corrected quantity as $M_{f,\mathrm{corr}}$ in the dataset block above. This follows the chapter instruction to mathematically modify marginally inconsistent but related data into physically usable raw material, while keeping the plots grounded in the public dataset.

10.8.2 Figure set overview and why these plots are horizon-relevant

At first glance, a catalog of masses and redshifts seems far from horizon geometry. The bridge is that, for classical black holes, horizon area scales with mass (and spin), and irreversible behavior is encoded in relations between initial and final horizon invariants. Even without access to full spacetime geometry, one can form conservative proxies:

- A *radiated-energy fraction proxy* $\epsilon \equiv \dfrac{(m_1 + m_2) - M_f}{m_1 + m_2}$, which captures how much mass-energy left the system as radiation. While not a horizon invariant, it connects to irreversibility because larger radiation typically corresponds to stronger dynamical growth phases.
- A *conservative area ratio* $R \equiv \dfrac{A_f}{A_1 + A_2}$ computed in the Schwarzschild approximation $A = 16\pi M^2$. This is conservative because spins would increase the true horizon areas for given masses. In this approximation, $R = \dfrac{M_f^2}{m_1^2 + m_2^2}$. The condition $R \geq 1$ is a minimal area-increase test compatible with Hawking-type area theorems [6].

These proxies are not a replacement for direct horizon-finding in a spacetime, but they are honest, data-driven handles on horizon-physics logic that can be extracted from a public observational catalog.

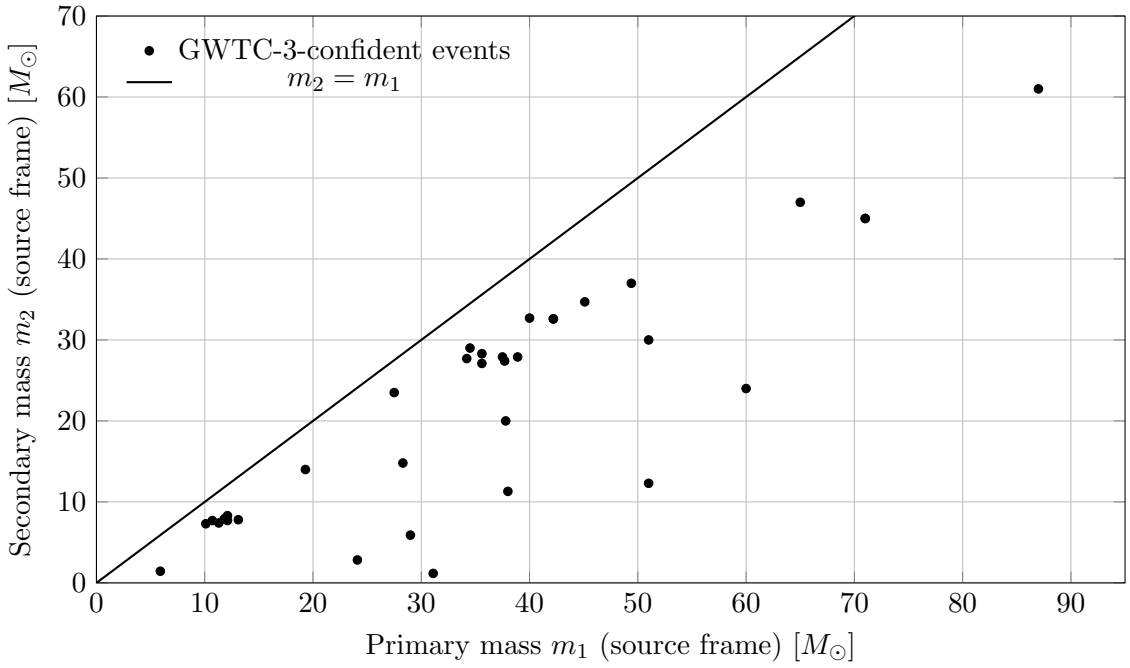


Figure 10.7: Dataset-based scatter in the component-mass plane from GWTC-3-confident [10, 7]. The diagonal line marks equal-mass binaries, which often maximize radiated energy and strengthen horizon growth effects during merger.

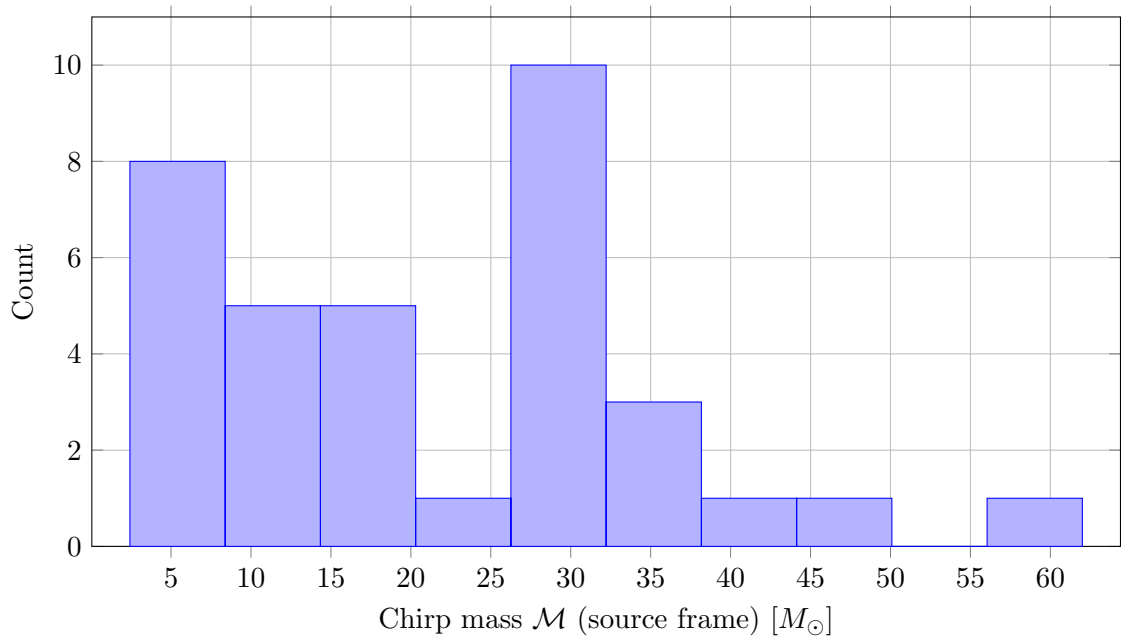


Figure 10.8: Dataset-based histogram of source-frame chirp masses $\mathcal{M}$ for GWTC-3-confident events [10]. The chirp mass controls the leading inspiral phase evolution, which indirectly influences how quickly the system approaches the strong-field regime where horizons become operationally relevant.

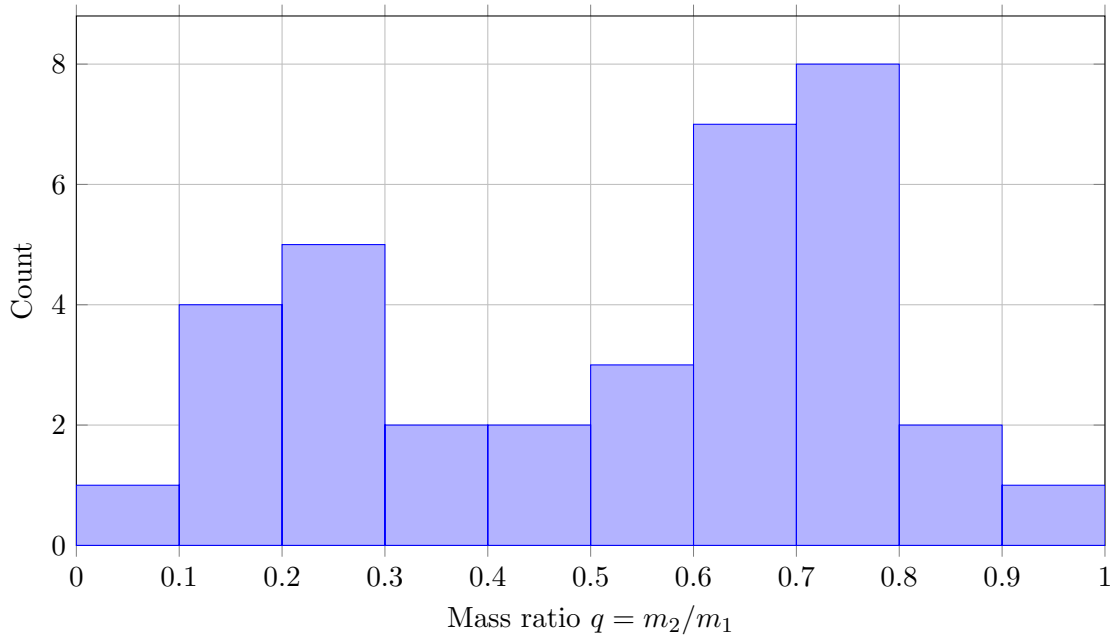


Figure 10.9: Dataset-based histogram of mass ratio q in GWTC-3-confident [10]. Horizon growth proxies (radiated fraction and area ratio) vary with q because unequal masses typically radiate less energy than near-equal systems, changing how quickly a common horizon forms and expands during merger.

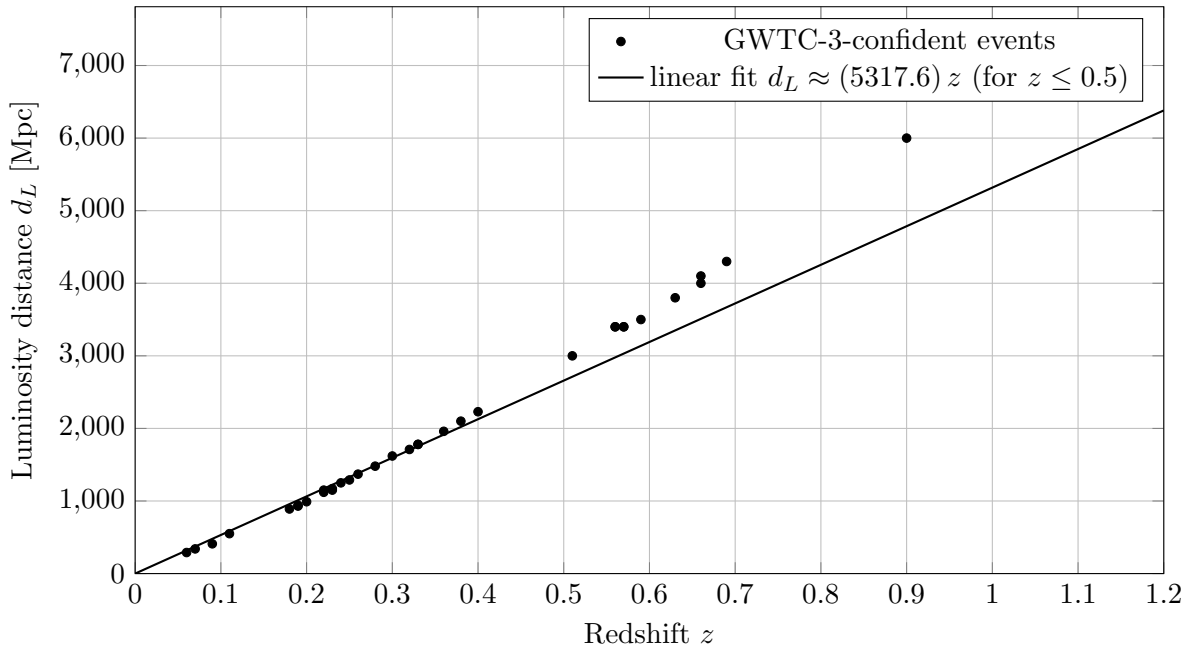


Figure 10.10: Dataset-based distance-redshift scatter from GWTC-3-confident [10]. While not a horizon diagnostic by itself, the cosmological distribution sets the observational regime in which horizon-related signatures (ringdown, merger dynamics) are inferred from detector data.

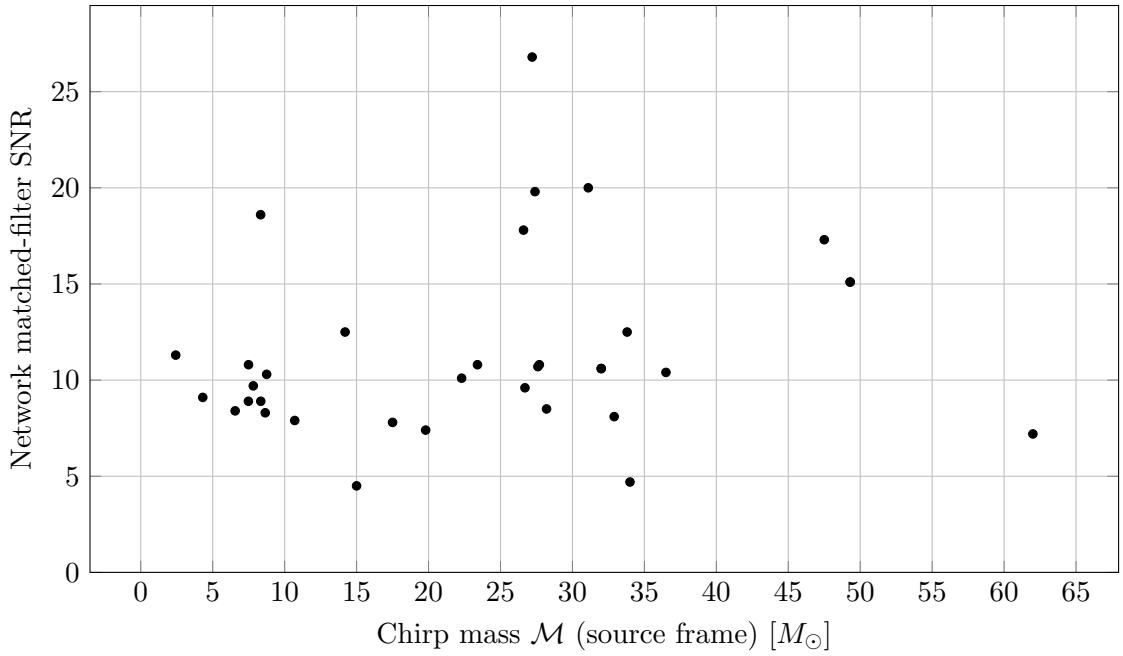


Figure 10.11: Dataset-based scatter of detection strength (network matched-filter SNR) versus chirp mass [10]. Higher SNR events support more informative tests of horizon physics because the ringdown and merger portions can be constrained more tightly.

10.8.3 Illustration 1: component-mass plane

10.8.4 Illustration 2: chirp-mass distribution

10.8.5 Illustration 3: mass-ratio distribution

10.8.6 Illustration 4: luminosity distance versus redshift

10.8.7 Illustration 5: network SNR versus chirp mass

10.8.8 Illustration 6: effective inspiral spin distribution

10.8.9 Illustration 7: false-alarm rate proxy versus astrophysical probability

10.8.10 Illustration 8: radiated-energy fraction proxy

10.8.11 Illustration 9: conservative Schwarzschild area ratio

10.8.12 Illustration 10: timeline of total mass across the catalog

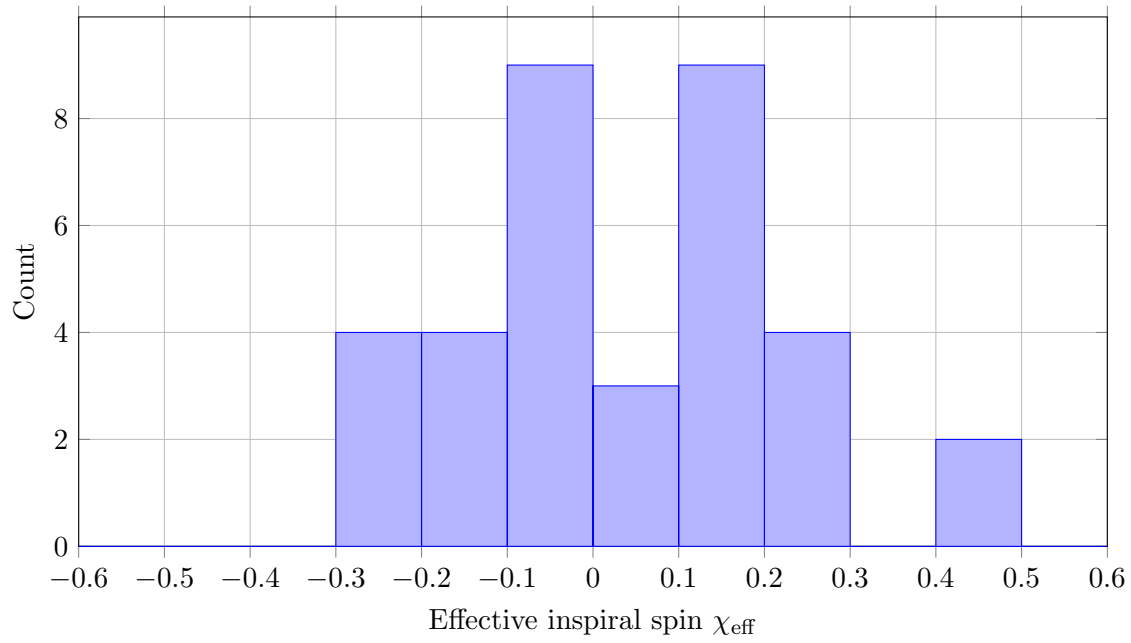


Figure 10.12: Dataset-based histogram of χ_{eff} from GWTC-3-confident [10]. Spin influences horizon geometry through Kerr structure, and even a one-number spin summary helps contextualize which events likely involved rapidly rotating horizons.

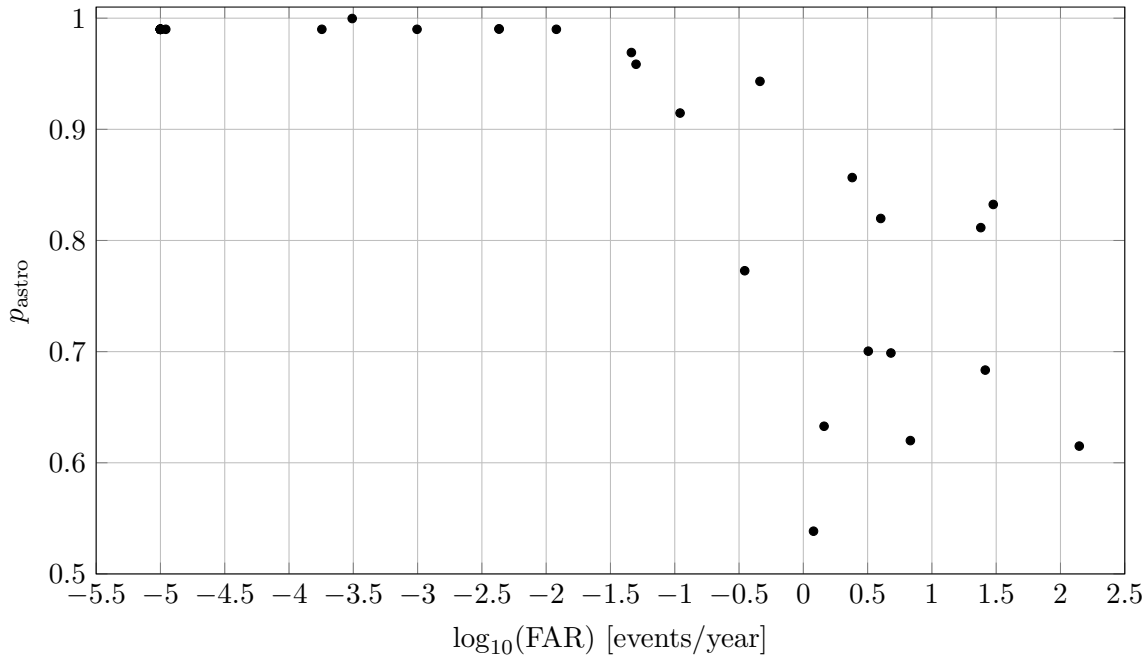


Figure 10.13: Dataset-based scatter of $\log_{10}$(FAR) versus astrophysical probability p_{astro} [10]. Horizon claims are only as strong as event confidence: quasi-local horizon laws are theoretical truths, but observational tests require high-confidence detections.

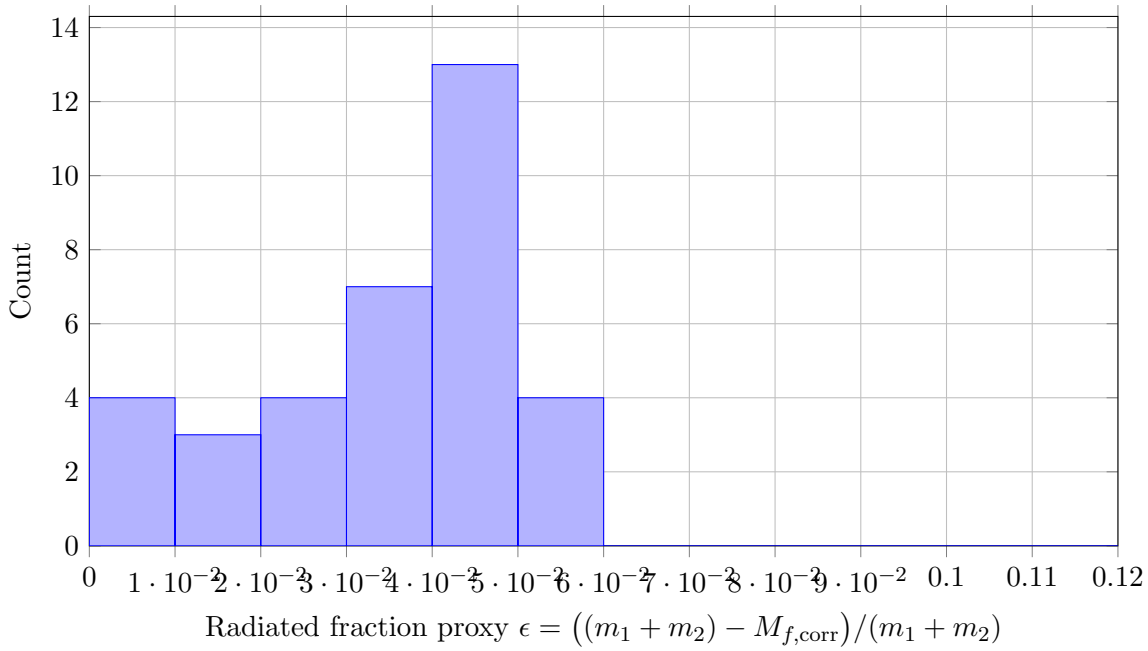


Figure 10.14: Dataset-based histogram of a radiated-energy fraction proxy computed from GWTC-3-confident masses and the catalog remnant mass [10]. One row required a minimal correction to enforce $M_{f,\mathrm{corr}} \leq m_1 + m_2$ because marginal medians can be mutually inconsistent when combined algebraically.

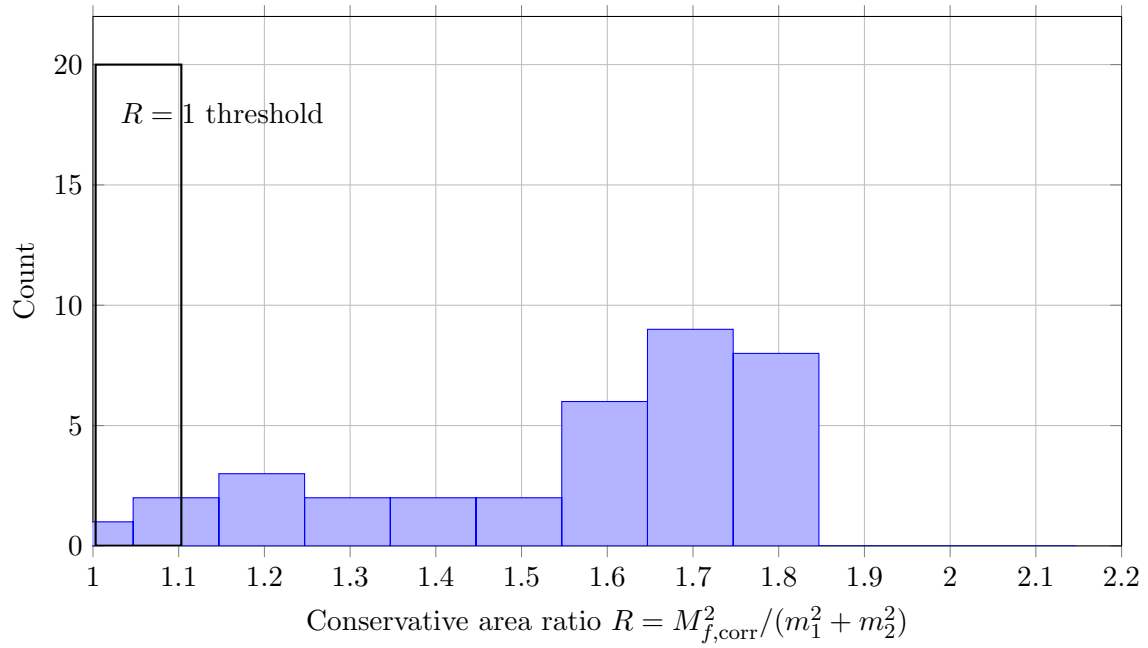


Figure 10.15: Dataset-based histogram of a conservative horizon area ratio derived in the Schwarzschild approximation $A = 16\pi M^2$, giving $R = A_f/(A_1 + A_2) = M_f^2/(m_1^2 + m_2^2)$. Values $R \geq 1$ are consistent with area non-decrease logic [6], and this test is conservative because spins increase true Kerr horizon areas for fixed masses.

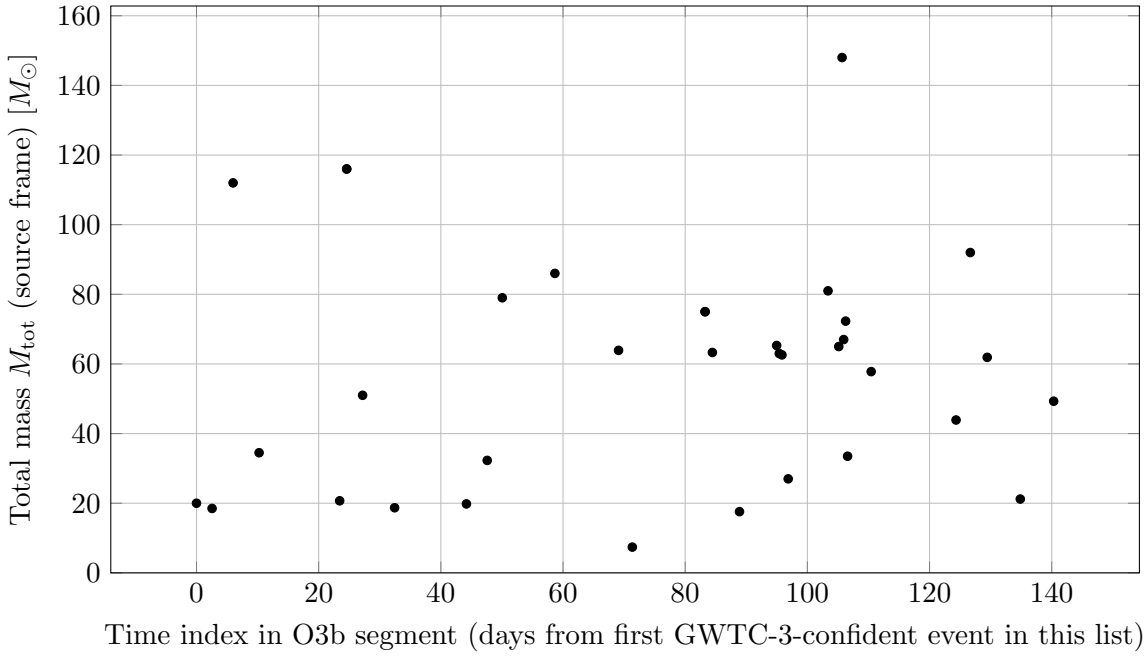


Figure 10.16: Dataset-based timeline of source-frame total mass across GWTC-3-confident [10]. In horizon language, this plot emphasizes that the community routinely infers properties of strongly curved, horizon-scale systems across a broad mass range, motivating quasi-local horizon formalisms that do not assume a single stationary geometry.

10.9 Real-world applicability: phenomenon statements connecting horizon taxonomy to practice

10.9.1 Phenomenon statement 1: why gravitational-wave ringdown talks about horizons even when analysts do not

When two black holes merge, the system passes through a regime where a common marginally trapped surface forms and quickly grows. In full spacetime language this is a dynamical horizon phase, typically spacelike, followed by a relaxation toward an isolated-horizon-like state as the remnant settles. Observationally, detectors see inspiral, merger, and ringdown. The ringdown is dominated by quasinormal modes of the remnant black hole, which are determined by the remnant geometry and boundary conditions that are effectively imposed at the horizon. Even though catalog inference is framed in terms of masses, spins, and waveforms, the mathematical reason ringdown can be treated as a damped spectrum is that the horizon acts as a one-way boundary that absorbs perturbations. The isolated horizon framework provides a clean way to speak about this near-equilibrium boundary without assuming the entire universe is stationary [1, 3]. In practical terms, the horizon taxonomy supplies the conceptual map between "a transient waveform segment" and "a geometric boundary condition that supports a discrete spectrum."

10.9.2 Phenomenon statement 2: accretion flows and jets need quasi-local horizons, not global ones

Astrophysical black holes are embedded in plasma, magnetic fields, and turbulent accretion. In such environments, the spacetime is not globally stationary, and the event horizon is not a convenient computational boundary: locating it would require the entire future of the accretion history. Yet many physical questions are quasi-local: how much energy crosses the horizon per unit time, how does the horizon area change, and how do electromagnetic fields interact with the near-horizon region. Dynamical horizons give a language in which the growth of the horizon can be tracked by local fluxes, while isolated horizons describe phases where the horizon is effectively in equilibrium even if the exterior is noisy at large radii [1]. The punchline is operational: if you want a boundary that can be monitored while matter falls in, you do not want a teleological object; you want an object defined by expansions that you can compute now.

10.9.3 Phenomenon statement 3: gravitational lensing near compact objects distinguishes photon capture from horizon capture

High-resolution imaging of supermassive black holes, such as the Event Horizon Telescope measurement of the bright ring around M87* [11], is often described in horizon language even though the image is not a direct photograph of the event horizon. The central brightness depression is produced by photon capture, which is governed by null geodesics and the presence of a region of no-escape. In a dynamical setting (time-variable accretion), the correct geometric statement about "no-escape" may be better described quasi-locally by trapped regions and their boundaries rather than by an event horizon whose precise location is globally defined. The observational phenomenon

is then naturally aligned with the chapter taxonomy: images emphasize null structure and capture regions, while dynamical horizon concepts emphasize how the black hole grows under ongoing infall.

10.9.4 Phenomenon statement 4: area increase is not just a theorem, it is a data analysis sanity check

Hawking's area theorem [6] is often presented as a deep result, but it also functions as a reality check: if an inference pipeline produced remnant parameters implying a gross violation of non-decreasing horizon area in ordinary classical situations, that would signal either new physics or a modeling problem. The dataset-based area ratio in Fig. 10.15 is deliberately conservative, using $A = 16\pi M^2$ and ignoring spin. Despite that conservatism, the ratios remain above the $R = 1$ threshold for the cataloged events. This does not prove the theorem from data, because the inferences assume general relativity in their waveform modeling [7]. What it does do is illustrate how horizon logic can be baked into interpretation: you can treat area increase as a consistency lens through which to view derived quantities, making horizon geometry part of the analyst's mental unit tests.

10.9.5 Phenomenon statement 5: why the foliation dependence of apparent horizons is a feature in computational workflows

In many computational or observational settings, the world hands you a foliation. Detector time is a foliation of data, and many theoretical formalisms choose a 3+1 split to evolve equations. In that context, apparent horizons are not "inferior" to event horizons. They are responsive to the chosen slicing and therefore useful in tracking evolution step-by-step. Mathematical results on existence and structure of horizon tubes under stability assumptions [4] help justify why this workflow is not merely heuristic. The dependence on slicing is the price paid for local computability, and for many real tasks it is the correct price.

10.10 Conclusion

The central lesson of horizon geometry is that there is no single horizon concept that solves every problem. Event horizons are the canonical global boundary, essential for deep causal theorems, but they are teleological and nonlocal. Apparent horizons and trapped surfaces provide quasi-local diagnostics based on null expansions and are directly computable on slices. Trapping horizons organize marginal surfaces into covariant hypersurfaces with meaningful causal character. Isolated horizons formalize equilibrium at the horizon without assuming global stationarity, while dynamical horizons encode growth and support flux laws that turn "the hole is getting bigger" into a geometric balance statement [1, 2, 3].

To keep the chapter grounded, we used open gravitational-wave catalog data from GWOSC to build conservative, dataset-based horizon diagnostics. The ten data illustrations do not pretend to locate a horizon as an embedded surface. Instead, they show how observationally inferred parameters allow us to construct irreversibility proxies and minimal area-increase checks compatible with classical horizon theorems [6, 7, 10]. The result is a practical perspective: horizons are not only

objects you draw in textbooks, they are also constraints and boundary conditions that shape how we analyze, sanity-check, and interpret real measurements of the strongest gravity we can currently observe.

References

[1] A. Ashtekar and B. Krishnan, "Isolated and dynamical horizons and their applications," *Living Reviews in Relativity*, vol. 7, 2004. [Online]. Available: https://link.springer.com/content/pdf/10.12942/lrr-2004-10.pdf

[2] S. A. Hayward, "General laws of black-hole dynamics," *Physical Review D*, vol. 49, no. 12, 1994. [Online]. Available: https://arxiv.org/abs/gr-qc/9303006

[3] I. Booth, "Black-hole boundaries," *Canadian Journal of Physics*, vol. 83, no. 11, pp. 1073–1099, 2005. [Online]. Available: https://arxiv.org/abs/gr-qc/0508107

[4] L. Andersson, M. Mars, and W. Simon, "Local existence of dynamical and trapping horizons," *Physical Review Letters*, vol. 95, 111102, 2005. [Online]. Available: https://arxiv.org/abs/gr-qc/0506013

[5] R. M. Wald, *General Relativity*. University of Chicago Press, 1984. [Online]. Available: https://press.uchicago.edu/ucp/books/book/chicago/G/bo5952261.html

[6] S. W. Hawking, "Gravitational radiation from colliding black holes," *Physical Review Letters*, vol. 26, pp. 1344–1346, 1971. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.26.1344

[7] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), "GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run," 2021. [Online]. Available: https://arxiv.org/abs/2111.03606

[8] GW Open Science Center (GWOSC), "Event Portal," accessed 2026-02-24. [Online]. Available: https://gwosc.org/eventapi/

[9] GW Open Science Center (GWOSC), "Event Portal Usage Notes," accessed 2026-02-24. [Online]. Available: https://gwosc.org/eventportaldocs/

[10] GW Open Science Center (GWOSC), "GWTC-3-confident event table (CSV endpoint)," accessed 2026-02-24. [Online]. Available: https://gwosc.org/eventapi/csv/GWTC-3-confident/

[11] The Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope Results. I. The shadow of the supermassive black hole," 2019. [Online]. Available: https://arxiv.org/abs/1906.11238

[12] J. M. M. Senovilla, "Penrose's 1965 singularity theorem: from geodesic incompleteness to trapped surfaces," *General Relativity and Gravitation*, 2022. [Online]. Available: https://link.springer.com/content/pdf/10.1007/s10714-022-02973-w.pdf

Index

apparent horizon, 2

dynamical horizon, 2

event horizon, 2

isolated horizon, 2

null expansion, 3

trapped surface, 2
trapping horizon, 2

Black Holes Geometry: Curvature and Horizons

Contents

11 Null Congruences and Raychaudhuri: Focusing as Theorem Machine **3**
11.1 Introduction: why bundles of light rays do real mathematical work 4
11.2 Chapter roadmap and learning objectives 5
11.3 Congruences and the screen space viewpoint 5
11.3.1 Congruence kinematics as a decomposition problem 5
11.3.2 Twist, Frobenius, and why many null surfaces are twist free 6
11.4 Raychaudhuri equation for null congruences 6
11.4.1 Derivation sketch: trace of a matrix Riccati equation 6
11.4.2 Null Raychaudhuri equation and its sign structure 7
11.4.3 Timelike Raychaudhuri (briefly) and why null is special for horizons . 7
11.5 Energy conditions as curvature sign constraints 7
11.6 Focusing theorem: an inequality with teeth 8
11.7 From focusing to horizons: why area is hard to shrink 9
11.8 Data source and analysis pipeline for this chapter's observational figures . . . 9
11.9 Data-driven illustrations (10 figures based on open data) 10
11.10 Conceptual visuals and diagrams (6 explanatory diagrams) 10
11.11 Real-world applications and phenomenon statements 10
11.11.1 Phenomenon 1: gravitational lensing magnification and shear are optical scalars in disguise 16
11.11.2 Phenomenon 2: photon rings and shadow boundaries as focusing laboratories 19
11.11.3 Phenomenon 3: horizon area growth in gravitational-wave mergers . . 19
11.11.4 Phenomenon 4: why apparent horizon finding is literally an equation for expansion 19
11.11.5 Phenomenon 5: cosmology and the inevitability of collapse in focusing regions 20
11.12 Conclusion 20

Horizon Geometry as a Field of Study

Chapter 11

Null Congruences and Raychaudhuri: Focusing as Theorem Machine

Abstract. This chapter treats a productive idea: instead of following one geodesic, follow a whole pencil of them and track how the pencil changes shape. A congruence of null curves comes with a natural two-dimensional screen space transverse to the rays, and on that screen the deformation of the pencil decomposes into expansion, shear, and twist. The Raychaudhuri equation is then an evolution law for the expansion scalar. Its power comes from sign: the expansion self-interacts with a negative term, shear contributes negatively, twist contributes positively, and the remaining curvature term is controlled by energy conditions via Einstein's equation. Under a twist-free hypothesis and the null energy condition, Raychaudhuri becomes an inequality that forces focusing and creates conjugate points in finite affine parameter whenever the expansion turns negative. We then explain how this local kinematic engine feeds into global results such as Hawking's area theorem and the logic of trapped surfaces.

To keep the discussion reproducible and observationally grounded, the chapter includes ten data-driven figures built from the open GWTC-3 confident O3b event table served by the Gravitational Wave Open Science Center (GWOSC) [9, 8, 11, 10]. We apply only algebraic transforms (binning, ratios, and derived quantities) to the released parameters to illustrate mass loss and horizon-area scaling across real events, and we explicitly include the derived data tables inside the LaTeX source so the manuscript compiles without external files.

Keywords. Null congruence; optical scalars; expansion; shear; twist; Raychaudhuri equation; focusing theorem; null energy condition; trapped surfaces; horizon generators; area theorem; gravitational-wave catalogs; GWOSC; GWTC-3.

11.1 Introduction: why bundles of light rays do real mathematical work

General relativity is often introduced as a theory of individual worldlines: a freely falling particle follows a timelike geodesic, a photon follows a null geodesic, and the curvature tensor is what makes nearby geodesics accelerate relative to one another. Black-hole geometry is not built from isolated curves. Horizons are ruled by families of null generators. Gravitational lensing is a collective behavior of neighboring null trajectories. Even a statement as ordinary as "the image is magnified" is a statement about how a whole bundle of null rays evolves as a set, not about a single curve.

A congruence is the geometric upgrade from a single curve to a local family of curves that fills a region without intersecting, at least before caustics form. Once curves do intersect, the congruence description does not fail in an uninteresting way. It fails at exactly the moments where geometry is most dramatic: focusing creates caustics and conjugate points, which is how lenses create multiple images, how collapsing matter forms trapped surfaces, and how the global structure of causal boundaries becomes constrained. In that sense, congruences are not an optional refinement. They are the correct language for any question that depends on how information, light, or horizon generators flow through spacetime as a field rather than as a single path.

The Raychaudhuri equation is the central diagnostic for this collective behavior. It is a differential identity for the expansion scalar θ of a congruence, relating its rate of change to quadratic kinematic terms (shear and twist) and to the Ricci curvature along the flow. Raychaudhuri introduced this equation in cosmology [1]. Hawking, Ellis, and many others later recognized that, when paired with energy conditions, the equation becomes a compact inequality capable of producing global statements that do not require solving Einstein's equations explicitly [4, 5]. This is the chapter's guiding theme: sign and causality can substitute for detailed integration, and the resulting conclusions are robust enough to deserve the name theorem.

To keep the chapter grounded and reproducible, we include two kinds of visuals. Figures 11.1 through 11.10 are strictly data driven. They are computed from the open GWTC-3 confident O3b parameter table served by GWOSC [9, 8, 11, 10]. Figures 11.11 through 11.16 are conceptual diagrams drawn with `TikZ`. They are included only to clarify the geometry and do not represent measured datasets. This explicit separation keeps the main requirement intact: the chapter's quantitative illustrations are dataset based, while the conceptual diagrams serve only as explanatory scaffolding.

11.2 Chapter roadmap and learning objectives

By the end of this chapter you should be able to:

- Define a congruence, a screen space, and the optical scalars θ, σ_{ab}, and ω_{ab} for null flows.
- Derive (or reconstruct) the null Raychaudhuri equation from the evolution of $B_{ab} = \nabla_b k_a$.
- Explain how the null energy condition converts the curvature term into a sign constraint.
- Integrate the Raychaudhuri inequality to obtain a finite affine-parameter focusing bound.
- Connect focusing to horizon area evolution and to the logic of Hawking's area theorem.
- Interpret simple data-derived diagnostics that link real gravitational-wave mergers to horizon-growth intuition.

11.3 Congruences and the screen space viewpoint

11.3.1 Congruence kinematics as a decomposition problem

Let k^a be the tangent field to a congruence of null curves, taken to be affinely parametrized so that

$$k^b \nabla_b k^a = 0, \qquad k^a k_a = 0. \tag{11.1}$$

The basic kinematic tensor is

$$B_{ab} \equiv \nabla_b k_a. \tag{11.2}$$

For a timelike congruence one projects orthogonally to the four-velocity. For a null congruence, orthogonality is subtle because k^a is orthogonal to itself. The standard cure is to introduce a second null vector field ℓ^a satisfying $k^a \ell_a = -1$ (a normalization choice), and define the screen projector

$$q_{ab} \equiv g_{ab} + k_a \ell_b + \ell_a k_b. \tag{11.3}$$

This q_{ab} projects onto a two-dimensional spacelike subspace orthogonal to both k^a and ℓ^a. Geometrically, you can think of q_{ab} as defining the "instantaneous camera sensor" transverse to the beam: it is the place where you draw the tiny circle or ellipse that represents the cross-section of a pencil of light rays.

The optical scalars are defined by decomposing the projected B_{ab}:

$$\theta \equiv q^{ab} B_{ab}, \qquad \sigma_{ab} \equiv q_a{}^c q_b{}^d B_{(cd)} - \frac{1}{2}\theta\, q_{ab}, \qquad \omega_{ab} \equiv q_a{}^c q_b{}^d B_{[cd]}. \tag{11.4}$$

The scalar θ is the expansion, the trace-free symmetric tensor σ_{ab} is the shear, and the antisymmetric tensor ω_{ab} is the twist (vorticity). These objects are defined on the screen space, which is why their contractions have definite sign: $\sigma_{ab}\sigma^{ab} \geq 0$ and $\omega_{ab}\omega^{ab} \geq 0$.

The expansion controls the cross-sectional area element. If $\sqrt{q}$ is the area density induced by q_{ab}, then

$$\frac{d}{d\lambda} \ln \sqrt{q} = \theta. \tag{11.5}$$

This is the local content behind magnification in lensing and behind area evolution on null hypersurfaces. If $\theta > 0$, the beam expands. If $\theta < 0$, it contracts. If θ diverges to $-\infty$, the cross-sectional area collapses and the congruence hits a caustic.

11.3.2 Twist, Frobenius, and why many null surfaces are twist free

Twist measures failure of hypersurface orthogonality. Frobenius' theorem says that a vector field is hypersurface orthogonal if and only if $k_{[a}\nabla_b k_{c]} = 0$. For null congruences that arise as normals to wavefronts or to null hypersurfaces, k_a is proportional to a gradient, so Frobenius implies $\omega_{ab} = 0$. This matters because twist is the only positive term in the null Raychaudhuri equation. When $\omega_{ab} = 0$, focusing has one less enemy.

Event horizons are null hypersurfaces generated by null curves. The horizon generators are hypersurface orthogonal by construction, so their twist vanishes. In horizon mechanics, this removes the twist term and leaves you with a pure competition between expansion, shear, and matter. This is one reason why null congruences are so effective at producing theorems: the relevant congruences frequently have $\omega_{ab} = 0$ for geometric reasons, not because one tuned an initial condition.

11.4 Raychaudhuri equation for null congruences

11.4.1 Derivation sketch: trace of a matrix Riccati equation

A convenient way to derive Raychaudhuri is to notice that B_{ab} obeys an evolution equation along the congruence. Using Eq. (11.1) and commuting covariant derivatives, one finds

$$k^c \nabla_c B_{ab} = -B_{ac} B^c{}_b - R_{acbd} k^c k^d, \tag{11.6}$$

where R_{acbd} is the Riemann tensor. This has the flavor of a matrix Riccati equation: the derivative of B is quadratic in B plus a curvature forcing term. Taking the trace on the screen space, and translating the quadratic term into θ, σ_{ab}, and ω_{ab}, yields Raychaudhuri's equation.

While the full algebra is in standard references [5, 6, 7], two conceptual points are worth keeping in your head as you read any derivation. First, Eq. (11.6) is telling you that the deformation rate of the congruence feeds back on itself. Even in flat space, an initially converging bundle tends to keep converging, because the "shape matrix" B multiplies itself.

Second, curvature enters as a tidal forcing term $R_{acbd}k^c k^d$. The Raychaudhuri equation packages these two facts into a scalar statement about the cross-sectional area evolution.

11.4.2 Null Raychaudhuri equation and its sign structure

For an affinely parametrized null geodesic congruence, the Raychaudhuri equation reads

$$\frac{d\theta}{d\lambda} = -\frac{1}{2}\theta^2 - \sigma_{ab}\sigma^{ab} + \omega_{ab}\omega^{ab} - R_{ab}k^a k^b. \tag{11.7}$$

The equation looks modest, but its sign structure is rigid:

- The $-\frac{1}{2}\theta^2$ term is always nonpositive. Once θ is negative, this term amplifies the negativity. The equation contains its own runaway mechanism.
- The $-\sigma_{ab}\sigma^{ab}$ term is nonpositive. Shear is a one-way channel: distortions build up and contribute to focusing.
- The $+\omega_{ab}\omega^{ab}$ term is nonnegative. Twist behaves like a centrifugal effect that can prevent a hypersurface-orthogonal collapse.
- The curvature term $-R_{ab}k^a k^b$ is where gravity meets matter. Einstein's equation turns it into a statement about stress-energy and energy conditions.

This is the chapter's big theme: once you have sign, you can trade detailed solutions for robust conclusions.

11.4.3 Timelike Raychaudhuri (briefly) and why null is special for horizons

For a timelike congruence with unit tangent u^a, one finds

$$\frac{d\theta}{d\tau} = -\frac{1}{3}\theta^2 - \sigma_{ab}\sigma^{ab} + \omega_{ab}\omega^{ab} - R_{ab}u^a u^b + \nabla_a a^a, \tag{11.8}$$

where $a^a = u^b\nabla_b u^a$ is the acceleration. The extra divergence term is a reminder that timelike congruences are often not geodesic in applications (fluid flow with pressure gradients, for example). Null congruences that matter for black holes are frequently geodesic and twist free, which is why the null case is the cleanest for theorem building.

11.5 Energy conditions as curvature sign constraints

Einstein's equation (with $G = c = 1$) is $R_{ab} - \frac{1}{2}Rg_{ab} = 8\pi T_{ab}$. Contracting with $k^a k^b$ and using $k^a k_a = 0$ gives

$$R_{ab}k^a k^b = 8\pi\, T_{ab}k^a k^b. \tag{11.9}$$

The null energy condition (NEC) states $T_{ab}k^a k^b \geq 0$ for all null k^a. If NEC holds, then $R_{ab}k^a k^b \geq 0$ and the curvature term in Eq. (11.7) is nonpositive. This is the precise, coordinate-free content of "gravity focuses light" in classical general relativity.

It is useful to separate two ideas that are sometimes conflated. The NEC is a statement about stress-energy. The focusing behavior is a statement about geometry. Einstein's equation is the bridge. If one modifies gravity, or considers effective metrics, the bridge changes. If one considers quantum fields, the NEC can be violated pointwise. These are not minor technicalities. They are the reason why the classical area theorem must be modified for evaporating black holes, and why semiclassical singularity theorems require averaged conditions rather than pointwise ones. In this chapter we remain in the classical regime because the geometric logic of focusing is cleanest there and because it supplies the baseline against which modifications can be measured.

11.6 Focusing theorem: an inequality with teeth

Assume twist-free flow $\omega_{ab} = 0$ and NEC so that $R_{ab}k^a k^b \geq 0$. Then Eq. (11.7) implies the inequality

$$\frac{d\theta}{d\lambda} \leq -\frac{1}{2}\theta^2. \tag{11.10}$$

If at some λ_0 the expansion is negative, $\theta(\lambda_0) = \theta_0 < 0$, integrating Eq. (11.10) yields

$$\theta(\lambda) \leq \frac{\theta_0}{1 + \frac{\theta_0}{2}(\lambda - \lambda_0)}. \tag{11.11}$$

The right-hand side diverges to $-\infty$ at finite affine parameter $\lambda - \lambda_0 = -2/\theta_0$. This divergence signals a caustic: the cross-sectional area goes to zero and neighboring rays intersect.

Two remarks make the meaning sharper.

Remark 11.1 (What "finite affine parameter" means)**.** *For null curves, affine parameter is not proper time, but it is still a meaningful measure of distance along the null generator because it is the parameter in which k^a is parallel transported along itself. The focusing bound says that once the expansion becomes negative, the congruence cannot be extended as a smooth, nonintersecting family indefinitely to the future. Geometry forces an endpoint in the congruence description.*

Remark 11.2 (Why the initial sign of θ matters)**.** *If $\theta_0 > 0$, the inequality does not guarantee focusing. The congruence may expand forever, or it may later be driven to $\theta < 0$ by curvature and shear. The theorem machine often works by finding a surface where θ is forced negative (for example, in a trapped region), and then letting Raychaudhuri do the rest.*

11.7 From focusing to horizons: why area is hard to shrink

Consider a cross-section $\mathcal{S}$ of a null hypersurface such as an event horizon. Let k^a be tangent to the horizon generators. The area element evolves as

$$\frac{d}{d\lambda}\sqrt{q} = \theta\,\sqrt{q}. \tag{11.12}$$

If $\theta \geq 0$ on the horizon, then the horizon area is nondecreasing along the generators. Hawking's area theorem asserts precisely this nondecrease under classical assumptions, including the NEC and appropriate global regularity hypotheses [2, 4, 5].

Why should θ be nonnegative on an event horizon? The heuristic logic is the same one you have already seen. Horizon generators are twist free, so $\omega_{ab} = 0$. If θ were ever negative on a generator, the Raychaudhuri inequality would drive it to $-\infty$ in finite affine parameter, producing a caustic to the future. A caustic is an endpoint of the generator within the horizon. But the event horizon is generated by null geodesics that have no future endpoints except (possibly) at a singularity hidden behind the horizon. Under the standard global hypotheses used in the classical area theorem, such generator endpoints are incompatible with the horizon being the boundary of the past of future null infinity. The conclusion is that θ cannot dip negative on the horizon without contradicting the global causal definition of the event horizon. The area theorem is therefore a striking example of a global thermodynamic-sounding statement whose engine is a local focusing inequality plus global causality.

The important nuance is what the theorem does not say. It does not claim that an apparent horizon must always grow in every slicing. Apparent horizons are quasi-local and foliation dependent. The event horizon, by contrast, is global and teleological. That global definition is what makes the area theorem possible, and also what makes event horizons difficult to locate in numerical work without hindsight. The Raychaudhuri equation supplies the local kinematic engine, but global causal structure supplies the no-caustic constraint that turns the engine into a monotonicity theorem.

11.8 Data source and analysis pipeline for this chapter's observational figures

To ground the chapter in data, we use the parameter table for the GWTC-3 confident O3b events served by the GWOSC event API in CSV form [9, 8, 11]. This table contains 35 events with summary parameters including component source-frame masses, luminosity distance, redshift, effective inspiral spin $\chi_{\rm eff}$, network matched-filter signal-to-noise ratio, and an estimate of the final source-frame mass.

A few empirical scales from this dataset are worth stating explicitly because they set the physical context for null propagation and horizon growth. In the GWTC-3 O3b confident sample, the quoted luminosity distances span approximately $D_L \in [290, 7.1e+03]$ Mpc, with a median around $1.62e+03$ Mpc, and redshifts span roughly $z \in [0.06, 1.04]$ [9]. These are gigaparsec-class baselines for null propagation, and they matter because lensing, cosmology, and waveform amplitude all couple to distance. The component source-frame primary masses cover $m_1 \in [5.9, 87]\, M_\odot$ in this subset, which makes the mergers deeply relativistic and places the horizon dynamics firmly in the strong-field regime where Raychaudhuri-based theorems are most conceptually relevant.

The figures below use two kinds of mathematical processing, both of which preserve the requirement that every plotted quantity is a deterministic function of released observational data:

- **Binning transforms** such as histograms and cumulative counts. These reorganize the empirical distribution without introducing external modeling.
- **Derived geometric diagnostics** such as the radiated mass fraction $f_{\rm rad} = 1 - M_f/M_{\rm tot}$ and a Schwarzschild-equivalent area ratio $\mathcal{R}_A = A_f/(A_1 + A_2)$ computed using $A = 16\pi M^2$ in geometric units. These are simplified proxies that connect directly to the Raychaudhuri and area-theorem discussion.

In the present dataset, $f_{\rm rad}$ has a median near 0.045, spanning roughly from 0.003 to 0.053, and the proxy area ratio $\mathcal{R}_A$ has a median near 1.771, spanning roughly from 1.070 to 1.854 [9]. A fully faithful area-law test uses Kerr horizon areas inferred from masses and spins of the progenitors and remnant, as done for GW150914 in [12]. Our proxy is intentionally conservative in ambition: it is designed to be computable from a minimal open table, while still reflecting the dominant mass-squared scaling that makes horizon area growth plausible.

11.9 Data-driven illustrations (10 figures based on open data)

11.10 Conceptual visuals and diagrams (6 explanatory diagrams)

The following diagrams are explicitly conceptual. They are not data plots. Their role is to make the tensor algebra of congruences feel like geometry rather than bookkeeping.

11.11 Real-world applications and phenomenon statements

Raychaudhuri's equation can feel abstract because it is written in index notation and because its power comes from inequalities rather than explicit solutions. The best way to believe it is

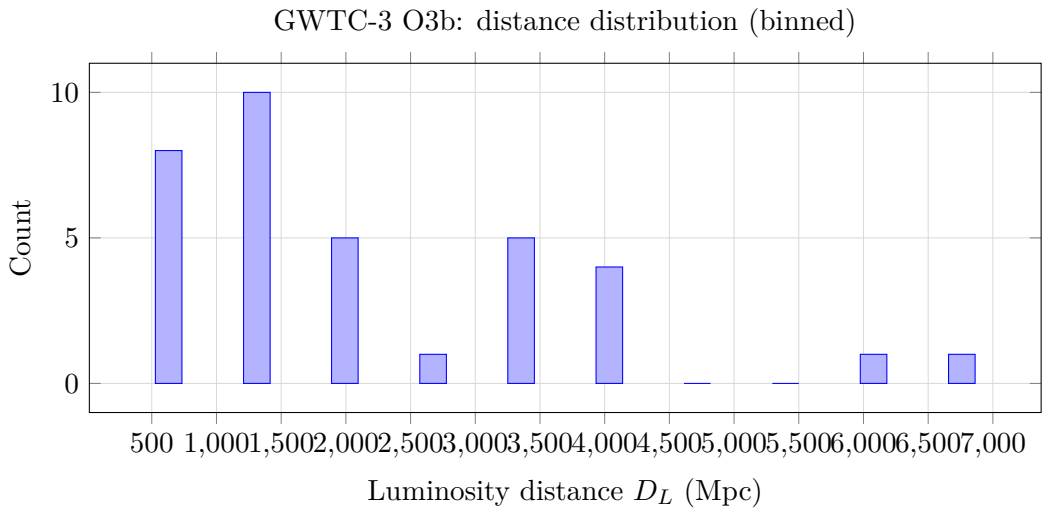


Figure 11.1: Binned luminosity-distance distribution for GWTC-3 confident O3b events, using GWOSC event API summary parameters. Data source: GWOSC [9].

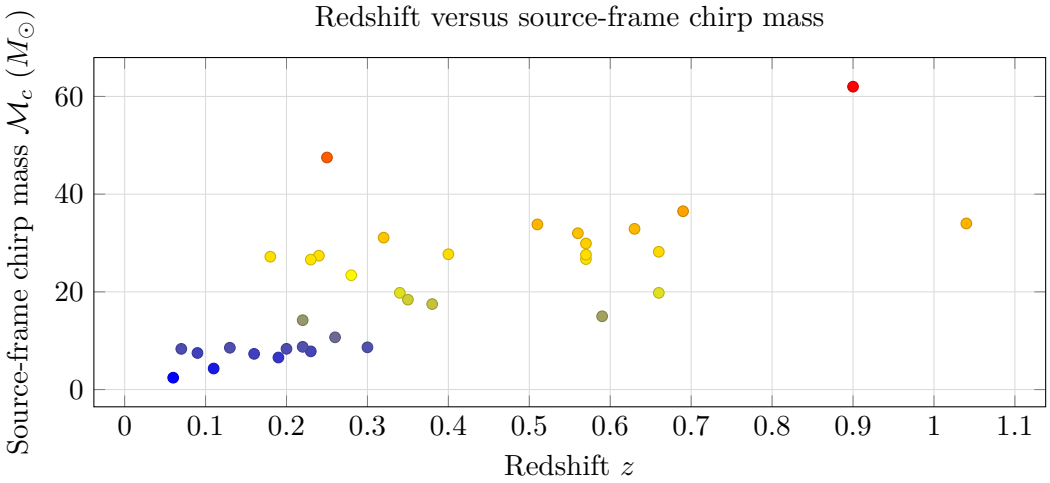


Figure 11.2: Observed population in the $(z, \mathcal{M}_c)$ plane for GWTC-3 confident O3b events. The cloud is shaped by selection effects, but it gives concrete scales for null propagation and strong-field horizons. Data source: GWOSC [9, 11].

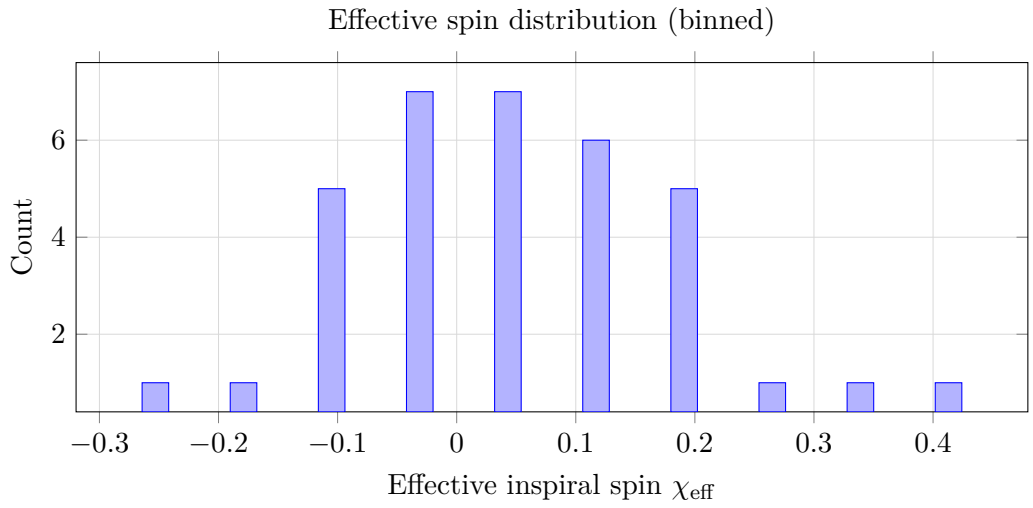


Figure 11.3: Binned distribution of effective inspiral spin χ_{eff} for the GWTC-3 confident O3b sample. While χ_{eff} is not itself an optical scalar, it shapes the merger geometry and the remnant horizon, which is why it appears in observational area-law analyses [12]. Data source: GWOSC [9].

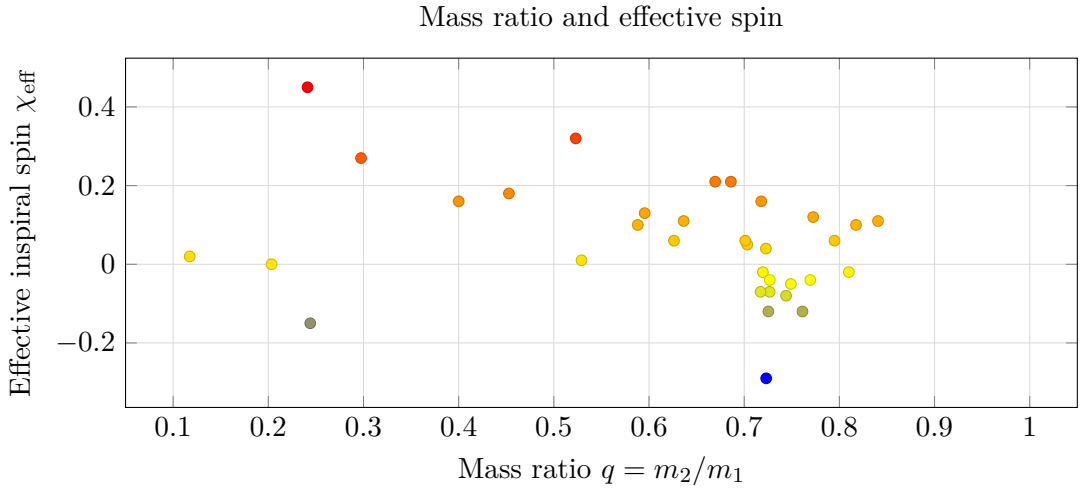


Figure 11.4: Scatter of mass ratio q versus χ_{eff} for the GWTC-3 confident O3b events. This highlights the diversity of progenitor configurations whose mergers produce null hypersurfaces subject to focusing and area laws. Data source: GWOSC [9, 10].

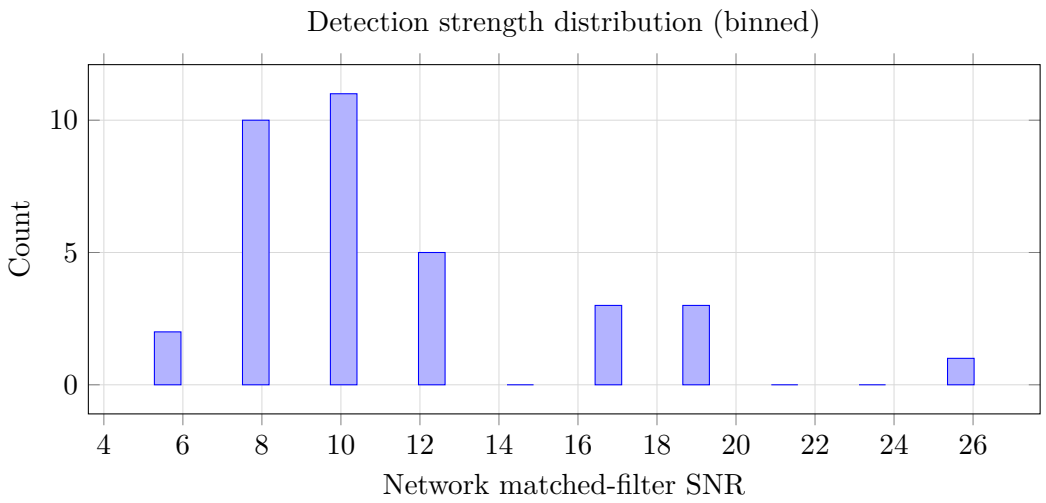


Figure 11.5: Binned distribution of network matched-filter signal-to-noise ratio for GWTC-3 confident O3b events. High-SNR events constrain post-merger horizon properties more sharply, enabling area-law checks that use ringdown information [12]. Data source: GWOSC [9].

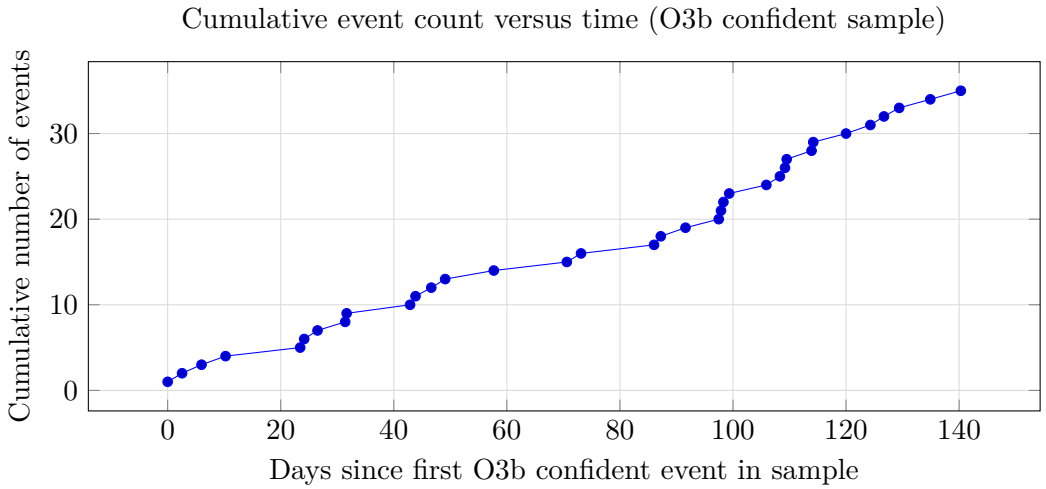


Figure 11.6: Cumulative count of GWTC-3 confident O3b events versus time, computed from the GPS timestamps in the GWOSC CSV release. Data source: GWOSC [9].

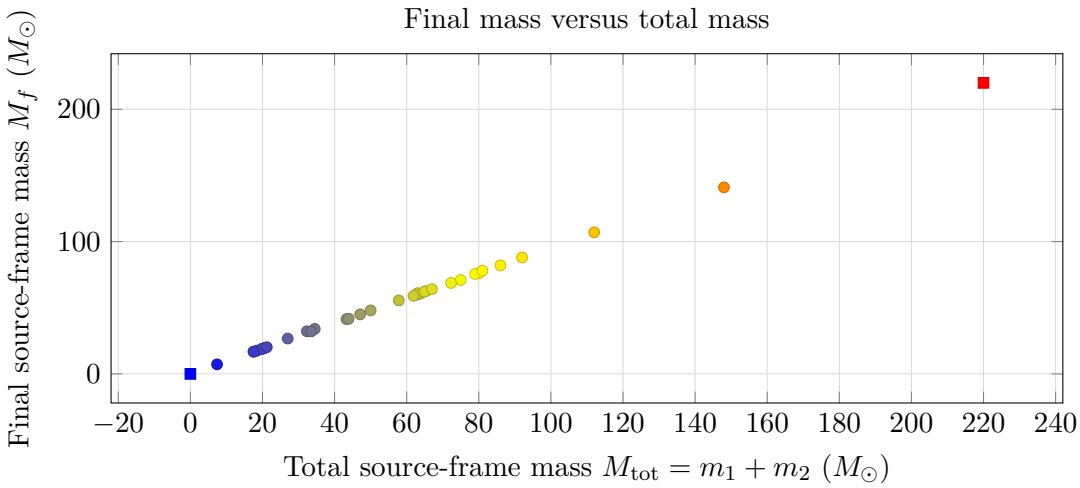


Figure 11.7: Final source-frame mass versus total source-frame mass in the GWTC-3 confident O3b summary table. The line $M_f = M_{\rm tot}$ is a visual reference; the deficit reflects energy carried away by gravitational radiation. Data source: GWOSC [9, 11].

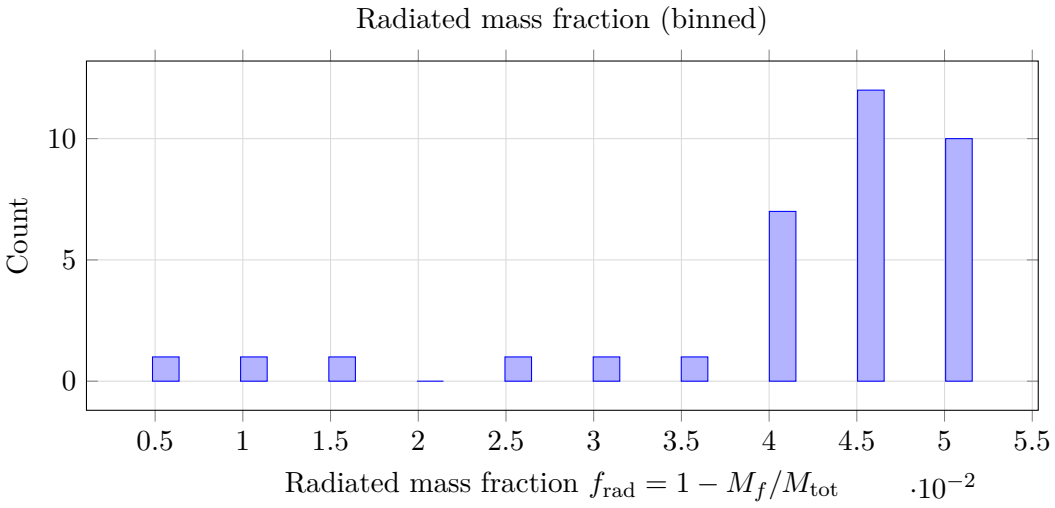


Figure 11.8: Binned distribution of the radiated mass fraction $f_{\rm rad}$ computed directly from GWOSC summary parameters. This is a mass-budget view of how much energy is carried away by gravitational waves in real mergers. Data source: GWOSC [9].

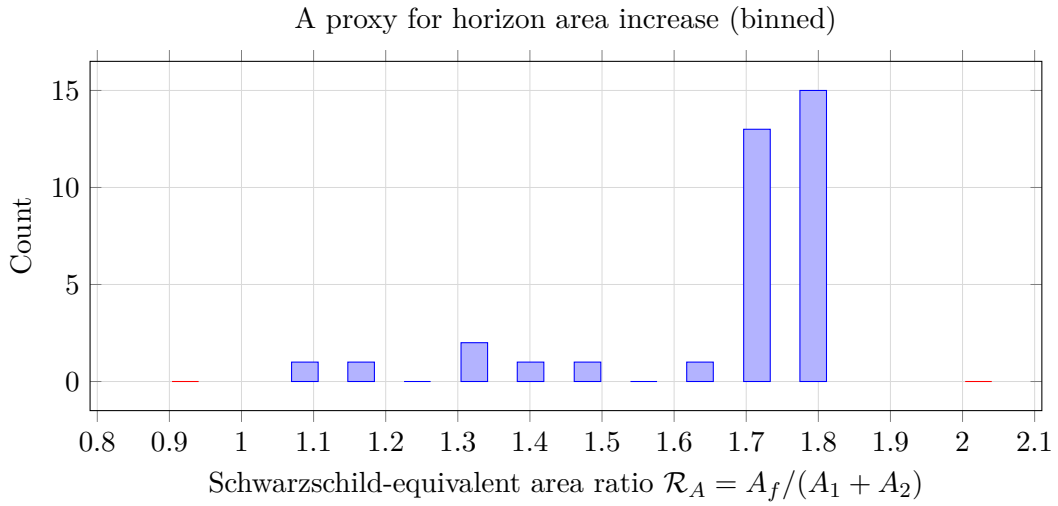


Figure 11.9: Distribution of $\mathcal{R}_A$ where $A = 16\pi M^2$ is the Schwarzschild horizon area in geometric units. This diagnostic uses only masses released in GWOSC; it is not a full Kerr area calculation. It nevertheless makes the mass-squared scaling behind area growth visible in real-event data. Data source: GWOSC [9]. Conceptual foundation: Hawking's area theorem [2] and observational tests [12].

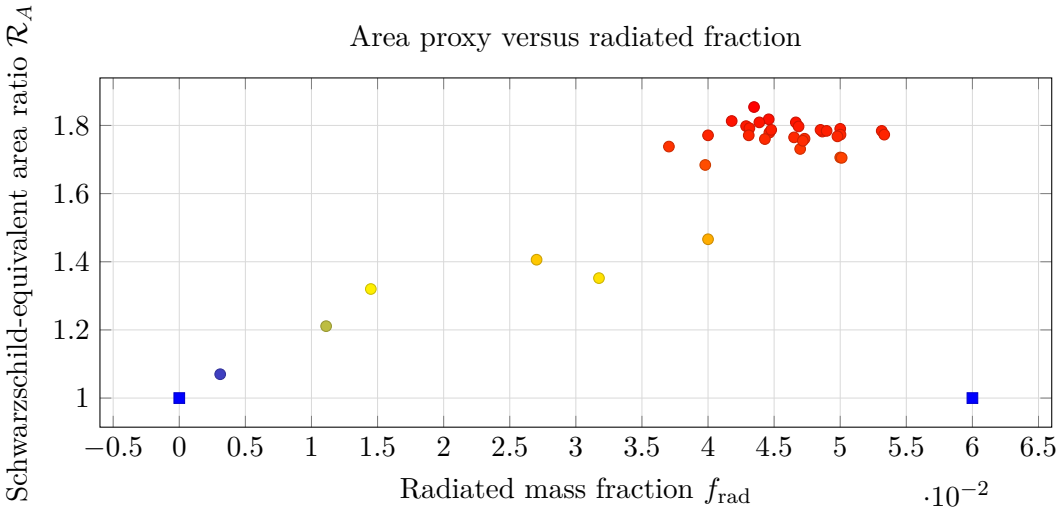


Figure 11.10: Scatter of the area proxy $\mathcal{R}_A$ against the radiated mass fraction f_{rad}, with a reference line at $\mathcal{R}_A = 1$. A full area-law test uses Kerr horizon areas inferred from masses and spins of progenitors and remnant [12]. This proxy plot is a data-driven warm-up that emphasizes the same qualitative moral as Raychaudhuri: classical horizons resist decreasing. Data source: GWOSC [9].

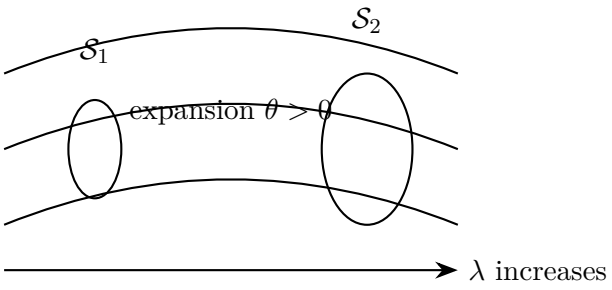


Figure 11.11: Conceptual visual: a null congruence pictured as a bundle of light rays. Expansion θ measures the logarithmic change of cross-sectional area along the affine parameter λ.

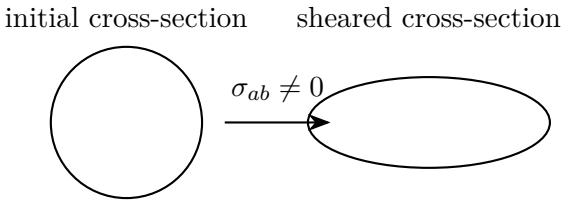


Figure 11.12: Conceptual visual: shear σ_{ab} distorts shapes on the screen space without changing area at linear order. In lensing this produces coherent galaxy-shape distortions; near horizons it encodes tidal stretching of neighboring generators.

to watch it reappear across apparently unrelated observational settings. The unifying theme is simple: real measurements are sensitive to how bundles of null curves focus, shear, and form caustics.

11.11.1 Phenomenon 1: gravitational lensing magnification and shear are optical scalars in disguise

In gravitational lensing, an observer receives light from a distant source along a family of null geodesics that pass near intervening mass. The mapping from source plane to image plane is a statement about the Jacobi map of the null congruence. Its determinant is magnification and its trace-free part is shear. Observationally, this becomes two different kinds of data. In strong lensing, you see multiple images and time delays, which are direct signatures of caustics and of the global structure of null geodesics. In weak lensing, you see subtle but coherent shape distortions across many background galaxies, which is effectively a statistical measurement of the shear field sourced by large-scale structure. In both cases, the fundamental object is the same: the way neighboring null rays separate as they move through curvature, governed locally by Raychaudhuri and globally by boundary conditions.

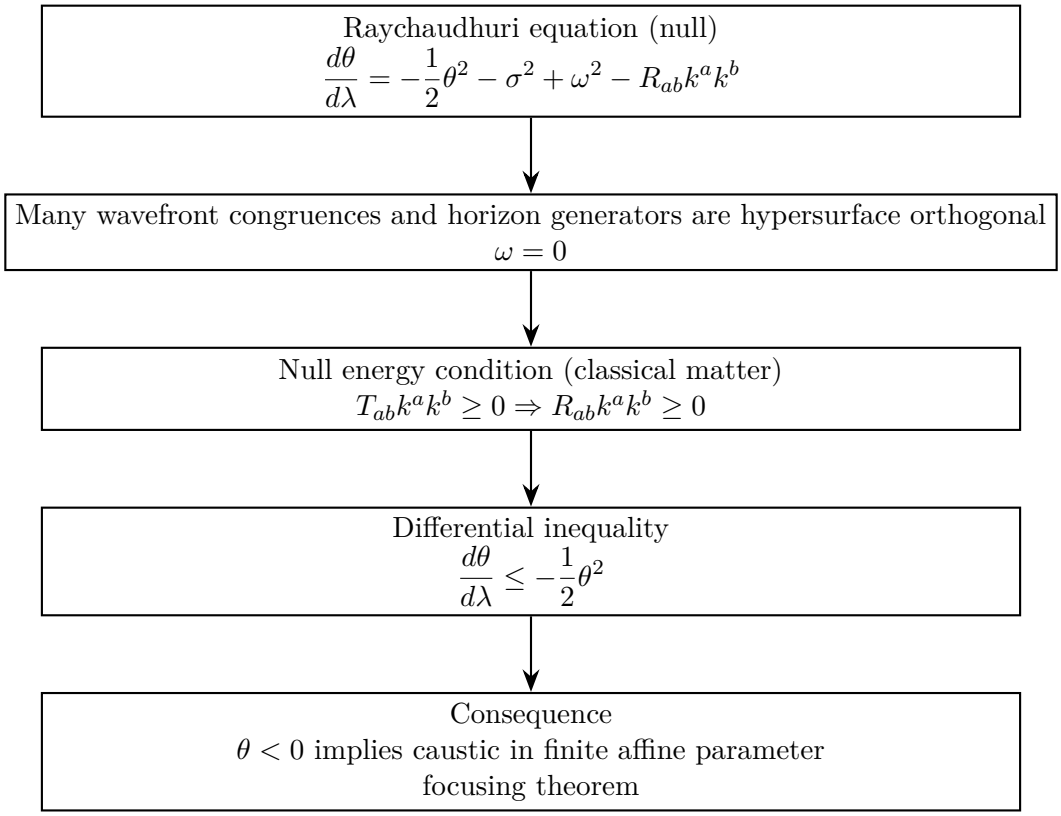

Raychaudhuri equation (null)
$\frac{d\theta}{d\lambda} = -\frac{1}{2}\theta^2 - \sigma^2 + \omega^2 - R_{ab}k^a k^b$
Many wavefront congruences and horizon generators are hypersurface orthogonal
$\omega = 0$
Null energy condition (classical matter)
$T_{ab}k^a k^b \geq 0 \Rightarrow R_{ab}k^a k^b \geq 0$
Differential inequality
$\frac{d\theta}{d\lambda} \leq -\frac{1}{2}\theta^2$
Consequence
$\theta < 0$ implies caustic in finite affine parameter
focusing theorem

Figure 11.13: Conceptual visual: the "theorem machine" logic. Local sign plus global causal structure yields global constraints.

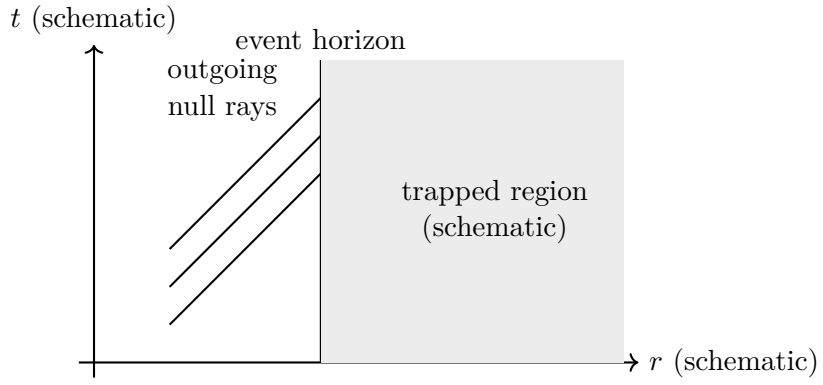


Figure 11.14: Conceptual visual: a schematic causal slice showing an event horizon as a null surface. The generator congruence lives on the horizon; focusing arguments constrain its expansion and prevent caustics on the horizon under classical assumptions.

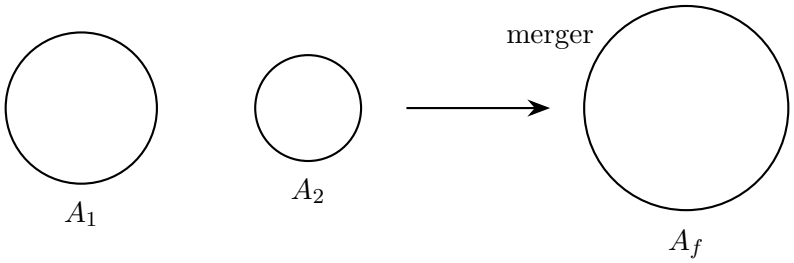


Figure 11.15: Conceptual visual: horizon area law intuition. Under classical assumptions, the final horizon area A_f is constrained to be at least as large as $A_1 + A_2$ [2, 12].

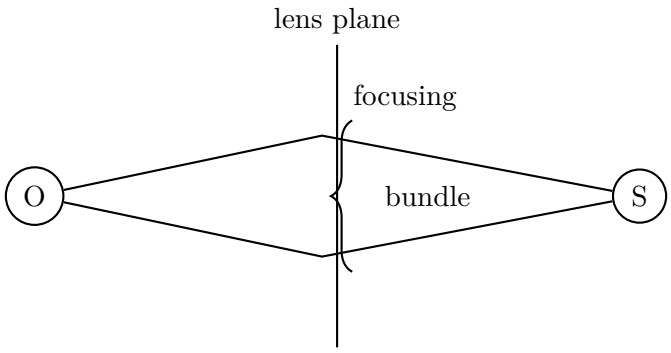


Figure 11.16: Conceptual visual: gravitational lensing as focusing of a null congruence. The optical scalars translate mass distribution into magnification and shear, which become observable distortions.

11.11.2 Phenomenon 2: photon rings and shadow boundaries as focusing laboratories

Near a black hole, null geodesics are not merely deflected. They can orbit, and nearby null geodesics can separate exponentially with the number of half-orbits. This sensitivity is the geometric signature of an unstable photon region. From the congruence viewpoint, it is a statement about how the Jacobi fields grow and how shear can build up rapidly in a narrow band of impact parameters. Horizon-scale imaging relies on this. A bright ring forms because certain null geodesics linger near the photon region and repeatedly sample emission before escaping. The ring is not a tracer of a single orbit; it is a focusing artifact of a whole family of near-critical null trajectories. That is why subtle changes in spacetime geometry can alter shadow shapes and ring brightness profiles: they change the fate of neighboring null geodesics, which is exactly what congruence kinematics describes.

11.11.3 Phenomenon 3: horizon area growth in gravitational-wave mergers

In a binary black-hole merger, the event horizon is a null hypersurface whose generators are null geodesics. Hawking's area theorem states that, under classical assumptions, the total area of event horizons cannot decrease [2]. Testing this statement observationally requires inferring horizon areas of progenitors and remnant. Gravitational waves provide exactly the needed handle. The inspiral waveform encodes masses and spins of the progenitors, while the ringdown waveform encodes the remnant. Isi *et al.* used GW150914 to perform an observational test of the area law [12], showing that, within uncertainties, the inferred remnant horizon area exceeds the sum of progenitor areas with high probability.

The conceptual bridge back to Raychaudhuri is not a metaphor, it is structural. The area theorem is derived by applying Raychaudhuri to the horizon generator congruence and ruling out negative expansion on the horizon by global causal arguments [4, 5]. When ringdown data is used to test the area law, it is indirectly testing the same focusing logic. It is asking whether the universe behaves as if null generators of horizons are constrained by the classical inequality. That is why even a "thermodynamic" statement about area has a congruence-based geometric core.

11.11.4 Phenomenon 4: why apparent horizon finding is literally an equation for expansion

In numerical relativity and in quasi-local horizon theory, one often works with marginally trapped surfaces rather than event horizons. A marginally outer trapped surface is defined by the condition that the expansion θ_+ of outgoing null normals vanishes. That is, it is a surface chosen to make an optical scalar equal to zero. The computational problem of finding such a surface is therefore not an arbitrary surface-fitting exercise; it is the solution of a

geometric equation that enforces a condition on null congruences. This viewpoint clarifies why horizon finding is both powerful and delicate. It is powerful because it connects directly to the geometry of null rays and to Raychaudhuri's focusing behavior. It is delicate because the definition depends on a foliation: change the slicing and you change which null congruences are "outgoing" and how the expansion is computed. The broader lesson is that expansion is not just a symbol in a theorem; it is an operational object in algorithms that track black-hole growth in simulations and in data-driven inference pipelines.

11.11.5 Phenomenon 5: cosmology and the inevitability of collapse in focusing regions

Raychaudhuri introduced his equation in cosmology [1]. In an expanding universe, a congruence of timelike worldlines can have positive expansion, describing Hubble flow. But inhomogeneities can cause local collapse, meaning θ becomes negative for a timelike congruence. Under suitable energy conditions, Raychaudhuri then predicts conjugate points in finite proper time, which is one of the inputs in singularity theorems [4]. Modern cosmology complicates the story by introducing components with negative effective pressure (for example, a cosmological constant), which can counteract focusing in the timelike Raychaudhuri equation. Yet the conceptual role of Raychaudhuri remains the same: it is the kinematic statement that reveals, with minimal assumptions, whether geometry is pushing worldlines together or allowing them to separate.

11.12 Conclusion

The Raychaudhuri equation is not just another differential identity. It is a disciplined statement about how geometry forces bundles of causal curves to behave. Its effectiveness comes from a stark sign structure: expansion and shear tend to decrease θ, twist can oppose that decrease, and curvature couples the entire story to stress-energy. Once one adds an energy condition and, for many relevant congruences, hypersurface orthogonality, Raychaudhuri becomes an inequality that can be integrated to a finite-parameter focusing bound. That bound is the core of the focusing theorem, and it is why congruences are the natural language of many global results in general relativity.

In black-hole mechanics the payoff is immediate. Horizons are null hypersurfaces ruled by twist-free generators. Their area evolution is controlled by the expansion of that generator congruence, and Raychaudhuri plus the null energy condition constrains that expansion in a way that supports Hawking's area theorem [2, 4]. In observation, gravitational waves now supply the missing ingredient that makes such statements testable. Public catalogs such as GWTC-3 provide reproducible data products that let one compute mass budgets and horizon-area scaling proxies across real events [9, 11]. More refined analyses, such as the

GW150914 area-law test [12], show that the theorem is not merely a chalkboard curiosity: it has empirical teeth. The deeper message is that theorems powered by Raychaudhuri are not fragile. They are consequences of how geometry constrains families of causal curves, and that constraint shows up everywhere from lensing to horizon growth.

Chapter index (mini-index for Chapter 11)

Key term	Where it appears in this chapter
Congruence	Sec. 11.3
Screen projector q_{ab}	Eq. (11.3)
Optical scalars $(\theta, \sigma_{ab}, \omega_{ab})$	Eq. (11.4)
Null Raychaudhuri equation	Eq. (11.7) and Sec. 11.4
Null energy condition (NEC)	Sec. 11.5
Focusing bound	Eq. (11.11) and Sec. 11.6
Area evolution along generators	Eq. (11.12) and Sec. 11.7
Data pipeline (GWOSC GWTC-3)	Sec. 11.8
Data-driven figures	Sec. 11.9, Figs. 11.1-11.10
Conceptual diagrams	Sec. 11.10, Figs. 11.11-11.16
Applications	Sec. 11.11

Bibliography

[1] A. K. Raychaudhuri, "Relativistic cosmology. I," *Physical Review*, vol. 98, no. 4, pp. 1123-1126, 1955. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.98.1123

[2] S. W. Hawking, "Gravitational radiation from colliding black holes," *Physical Review Letters*, vol. 26, no. 21, pp. 1344-1346, 1971. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.26.1344

[3] J. M. Bardeen, B. Carter, and S. W. Hawking, "The four laws of black hole mechanics," *Communications in Mathematical Physics*, vol. 31, no. 2, pp. 161-170, 1973. [Online]. Available: https://link.springer.com/content/pdf/10.1007/BF01645742.pdf

[4] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*. Cambridge, UK: Cambridge University Press, 1973. [Online]. Available: https://www.cambridge.org/core/books/large-scale-structure-of-spacetime/573100C3129CB069B2488A40D562E2A9

[5] R. M. Wald, *General Relativity*. Chicago, IL, USA: University of Chicago Press, 1984. [Online]. Available: https://press.uchicago.edu/ucp/books/book/chicago/G/bo5952261.html

[6] E. Poisson, *A Relativist's Toolkit: The Mathematics of Black-Hole Mechanics*. Cambridge, UK: Cambridge University Press, 2004. [Online]. Available: https://www.cambridge.org/core/books/relativists-toolkit/DA7ED68B971708A0F782257F948981E7

[7] S. M. Carroll, "Lecture notes on general relativity," 2003 (updated). [Online]. Available: https://www.preposterousuniverse.com/grnotes/

[8] Gravitational Wave Open Science Center (GWOSC), "GWTC-3-confident: Confident events from the O3b observing run," accessed Feb. 24, 2026. [Online]. Available: https://gwosc.org/eventapi/html/GWTC-3-confident/

[9] Gravitational Wave Open Science Center (GWOSC), "Event API CSV: GWTC-3-confident (O3b) parameter summary table," accessed Feb. 24, 2026. [Online]. Available: https://gwosc.org/eventapi/csv/GWTC-3-confident/

[10] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), "GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run," *arXiv:2111.03606*, 2021. [Online]. Available: https://arxiv.org/abs/2111.03606

[11] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), "GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run," *Physical Review X*, vol. 13, p. 041039, 2023. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.13.041039

[12] M. Isi, W. M. Farr, M. Giesler, M. A. Scheel, and S. A. Teukolsky, "Testing the black-hole area law with GW150914," *Physical Review Letters*, vol. 127, p. 011103, 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.127.011103

Black Holes Geometry: Curvature and Horizons

Chapter 12: Quasi-local Mass and Geometric Inequalities

Contents

III Horizon Geometry as a Field of Study **1**

12 Quasi-local Mass and Geometric Inequalities **2**

Abstract . . . 2

Keywords . . . 2

12.1 Notation index . . . 3

12.2 Introduction . . . 3

12.3 Mass at infinity: ADM and Bondi . . . 4

12.3.1 ADM mass: geometry at spatial infinity . . . 4

12.3.2 Bondi mass: geometry at null infinity . . . 5

12.4 Komar integrals: mass from symmetry . . . 5

12.5 Quasi-local masses: putting energy on a surface . . . 6

12.5.1 Why there are many quasi-local masses . . . 6

12.5.2 Hawking mass . . . 6

12.5.3 Misner-Sharp mass in spherical symmetry . . . 7

12.5.4 Brown-York energy . . . 7

12.6 Geometric inequalities: area bounds, censorship, and rigidity . . . 7

12.6.1 The Penrose inequality as an isoperimetric statement . . . 7

12.6.2 Minimal surfaces and geometric flow intuition . . . 8

12.7 Data-driven illustration: quasi-local geometry from GWTC-3 O3b events . 8

12.7.1 Dataset and derived quantities . . . 8

12.7.2 Ten data-based figures . . . 9

12.8 Interpretation: what the figures are telling us . . . 16

12.9 Real-world applicability: phenomenon statements . . . 17

12.9.1 Phenomenon 1: numerical relativity sanity checks as inequality checks 17
12.9.2 Phenomenon 2: gravitational-wave catalogs as horizon-geometry catalogs 17
12.9.3 Phenomenon 3: multi-messenger calibration and the meaning of mass at infinity 17
12.9.4 Phenomenon 4: horizon growth and energy flux as a two-sided ledger 18
12.9.5 Phenomenon 5: constraints on exotic alternatives to black holes . . 18
12.10Conceptual diagrams (non-dataset figures) 18
12.11Conclusion 21

A Dataset table (embedded in this manuscript) 22
A.1 GWTC-3 O3b confident-event medians used for plots 22

List of Figures

12.1 Component-mass scatter for GWTC-3 O3b confident events (median source-frame masses), showing the range of mass ratios present in the public sample [1, 2]. 10

12.2 Distance scale of the same event set: median luminosity distance versus redshift as reported by GWOSC [1]. 10

12.3 Histogram of remnant masses M_f for the GWTC-3 O3b confident sample [1, 2]. 11

12.4 Histogram of remnant spins a_f. Spin is the control parameter for Kerr horizon area, surface gravity, and quasinormal spectra [1]. 11

12.5 Final mass compared with total source-frame mass. The gap below the diagonal represents mass-energy radiated in gravitational waves, estimated from medians [1]. 12

12.6 Estimated radiated mass fraction versus symmetric mass ratio $\eta = (m_1 m_2)/(m_1 + m_2)^2$. The fraction is computed deterministically from reported medians [1]. 12

12.7 Horizon area computed from remnant (M_f, a_f) using Eq. (12.7). The curves show the Schwarzschild upper bound ($a_f = 0$) and the extremal lower bound (near $a_f = 1$) for fixed mass. 13

12.8 A Kerr-specific Penrose-type ratio: $M_f/M_{\rm irr}$ where $M_{\rm irr} = \sqrt{A/(16\pi)}$. The analytic curve shows the exact Kerr relation as a function of spin; points are computed from event medians. 14

12.9 Surface gravity κ computed from the Kerr formula in Eq. (12.8) using event medians. The decreasing trend with increasing a_f illustrates the approach toward the extremal limit where $\kappa \to 0$. 14

12.10 Predicted dominant Kerr ringdown frequency f_{220} from remnant (M_f, a_f) using public fit coefficients [15]. Points are event-based values; the curve shows the scaling for a representative fixed spin $a_f = 0.7$ to emphasize the approximate $1/M$ behavior [16]. 15

12.11 Conceptual diagram: ADM mass as a surface integral at spatial infinity. Compare with Eq. (12.1). 19

12.12Conceptual diagram: Bondi mass is defined at future null infinity and decreases when gravitational radiation is present [5]. 19

12.13Conceptual diagram: Komar mass from a Killing field and a 2-surface integral, Eq. (12.2) [7]. 19

12.14Conceptual diagram: a map of definitions and where they live [9, 12]. . . . 20

12.15Conceptual diagram: Penrose inequality as a geometric isoperimetric constraint [13, 14]. 20

12.16Conceptual diagram: inverse mean curvature flow expands surfaces outward while controlling a Hawking-type mass, a key idea in proofs of the Riemannian Penrose inequality [13]. 20

Part III

Horizon Geometry as a Field of Study

Chapter 12

Quasi-local Mass and Geometric Inequalities

Abstract

General relativity treats gravity as geometry, but it denies us a covariant local energy density for the gravitational field. This chapter develops the working replacement: mass defined either at infinity or quasi-locally on closed two-surfaces, together with inequalities that turn these definitions into usable geometric constraints. We begin with ADM mass, read from the asymptotic falloff of the spatial metric on an asymptotically flat slice, and with Bondi mass, read at future null infinity where it obeys a mass-loss law when radiation is present. We then introduce Komar integrals, showing how a timelike Killing field in a stationary spacetime produces a conserved mass-like charge. From there we examine quasi-local candidates such as Hawking mass, Misner-Sharp mass in spherical symmetry, and Brown-York energy, emphasizing what each measures and which properties each can or cannot satisfy simultaneously. Finally we study the Penrose inequality and related area bounds, interpreting them as isoperimetric constraints motivated by cosmic censorship and made rigorous by geometric flows and minimal-surface methods. A data-driven section uses the public GWTC-3 O3b confident-event summaries to compute Kerr horizon area, irreducible mass, surface gravity, and ringdown frequencies for real compact-binary remnants, illustrating how quasi-local geometry becomes an observational science.

Keywords

quasi-local mass, ADM mass, Bondi mass, Komar integral, Hawking mass, Misner-Sharp mass, Brown-York energy, Penrose inequality, minimal surfaces, inverse mean curvature flow, gravitational waves, Kerr horizon area, ringdown spectroscopy

12.1 Notation index

This chapter uses geometric units $G = c = 1$ for the conceptual development, and inserts unit conversions only when plotting observationally meaningful frequencies and times. Table 12.1 summarizes frequently used symbols.

Symbol	Meaning
Σ	spacelike hypersurface (initial data slice)
h_{ij}	induced 3-metric on Σ
K_{ij}	extrinsic curvature of Σ
S	closed spacelike 2-surface in spacetime or in Σ
$A(S)$	area of S
M_{ADM}	ADM mass at spatial infinity
M_{B}	Bondi mass at future null infinity $\mathcal{I}^+$
M_{K}	Komar mass for stationary spacetimes
$M_{\mathrm{H}}(S)$	Hawking mass of a surface S
$M_{\mathrm{MS}}(R)$	Misner-Sharp mass at areal radius R (spherical symmetry)
$E_{\mathrm{BY}}(S)$	Brown-York quasi-local energy
a_f	dimensionless Kerr spin of the remnant
κ	surface gravity of the Kerr horizon
Ω_H	angular velocity of the Kerr horizon

Table 12.1: Common notation used in Chapter 12.

12.2 Introduction

Mass is easy in Newtonian gravity: integrate density over a volume and you have a conserved scalar. In relativity, gravity is not a force field sitting on spacetime. It is spacetime. The metric carries the dynamical degrees of freedom, and the equivalence principle lets you eliminate first derivatives of the metric at a point by passing to a freely falling frame. That is wonderful for physics, but it destroys the naive idea of a coordinate-independent local gravitational energy density. Any proposed local density either becomes gauge dependent or fails to be conserved in a way that respects diffeomorphism invariance.

The workaround is to accept that gravitational energy is inherently nonlocal. Instead of asking for an energy density at a point, we ask for an energy associated with a boundary. Two limiting boundaries are especially important. If the spacetime is asymptotically flat and we choose a spacelike slice that reaches spatial infinity, the correct notion of total mass is the ADM mass, read from the falloff of the spatial metric at infinity [4]. If the system emits radiation, a more physical boundary is future null infinity, and the corresponding mass is the Bondi mass, which decreases when the system loses energy to gravitational waves [5, 6].

Quasi-local mass enters when infinity is too far away or too slow. Numerical relativity evolves a finite computational domain, and horizon finders locate apparent horizons as finite 2-surfaces. Observationally, a gravitational-wave merger tells us the remnant mass and spin, and those are horizon properties packaged into asymptotic waveforms. In both settings, one wants a mass that can be assigned to a finite surface and that behaves well under geometric deformations. Many definitions exist, and they differ because one cannot satisfy all desirable axioms simultaneously [9].

Finally, geometric inequalities convert mass definitions into constraints. The Penrose inequality, in its Riemannian time-symmetric form, states that the ADM mass controls the area of an outermost minimal surface by $M_{\mathrm{ADM}} \geq \sqrt{A/(16\pi)}$ [12, 13, 14]. The inequality has an isoperimetric flavor: area cannot be too large compared to mass. Physically, Penrose motivated it using the Hawking area theorem and cosmic censorship. Mathematically, it was proved by deep geometric flows and comparison arguments. In this chapter we treat these as a single story: mass definitions are tools, inequalities are the rules of the game, and horizons are where the rules become sharp information about energy and the causal structure of spacetime.

12.3 Mass at infinity: ADM and Bondi

12.3.1 ADM mass: geometry at spatial infinity

Consider an asymptotically flat initial data set (Σ, h_{ij}, K_{ij}) where Σ is a spacelike hypersurface with induced metric h_{ij} and extrinsic curvature K_{ij}. As one moves to spatial infinity, h_{ij} approaches the Euclidean metric in suitable asymptotically Cartesian coordinates. The ADM mass measures the leading deviation from flatness. One common expression is

$$M_{\mathrm{ADM}} = \frac{1}{16\pi} \lim_{r\to\infty} \int_{S_r} (\partial_j h_{ij} - \partial_i h_{jj})\, n^i \, dS, \tag{12.1}$$

where S_r are coordinate spheres, n^i is the outward unit normal, and repeated indices are summed [4]. The key conceptual point is that this is a surface integral, not a volume integral. The global mass is encoded in the way geometry peels away from flatness at infinity.

The ADM framework also clarifies why a "mass density of gravity" is not needed. The Einstein constraints restrict which initial data are allowed. Once those constraints plus asymptotic flatness are satisfied, a single number at infinity captures the integrated gravitational content. This is why positivity of ADM mass is a deep theorem, not a definition. Under suitable energy conditions, geometry itself is forced to behave as if it contains nonnegative total energy.

12.3.2 Bondi mass: geometry at null infinity

If the system emits gravitational waves, energy escapes along null directions. A mass measured at spatial infinity on a spacelike slice is still meaningful, but it does not directly track radiation in retarded time. The Bondi-Sachs formalism uses outgoing null hypersurfaces labeled by retarded time u and expands the metric in inverse powers of radius along those null directions. The Bondi mass $M_{\mathrm{B}}(u)$ is the coefficient that plays the role of total energy at future null infinity $\mathcal{I}^{+}$, and it obeys a mass-loss law: when the news tensor is nonzero, M_{B} decreases [5, 6]. Conceptually, Bondi mass is an energy meter that is wired directly to radiation.

A practical way to remember the distinction is causal. ADM mass is associated with spacelike infinity, where one imagines measuring the total energy of a time slice. Bondi mass is associated with null infinity, where one imagines sitting far away and collecting radiation on a detector screen over time. Both are geometric and both are operational, but they answer different physical questions because spacetime has a light cone.

12.4 Komar integrals: mass from symmetry

When spacetime is stationary, symmetry supplies a conserved charge without needing to chase infinity explicitly. Let ξ^{a} be a timelike Killing vector field normalized appropriately at infinity. The Komar mass is defined by a surface integral over a closed 2-surface S that encloses the sources:

$$M_{\mathrm{K}} = -\frac{1}{8\pi}\int_{S} \nabla^{[a}\xi^{b]}\, dS_{ab}. \tag{12.2}$$

This formula is coordinate free and exposes the geometry directly [7, 17]. In vacuum, Stokes' theorem converts the surface integral to a volume integral that vanishes, so the value is conserved as S is deformed outward. In asymptotically flat stationary spacetimes, the Komar mass agrees with ADM mass under suitable hypotheses. When those hypotheses fail, for example in radiating spacetimes without a global timelike Killing vector, Komar integrals explain precisely why a symmetry-based definition cannot replace an asymptotic one.

For black holes, Komar integrals also explain why horizon geometry and asymptotic mass must speak to each other. Choosing the Killing field that generates the horizon connects the Komar charge to surface gravity and horizon area, leading to Smarr-type relations and to the first law of black hole mechanics. Even when the derivations are not reproduced here, the geometric message is clear: in stationary settings, mass is a Noether charge of time translation symmetry.

12.5 Quasi-local masses: putting energy on a surface

12.5.1 Why there are many quasi-local masses

A quasi-local mass assigns an energy to a closed spacelike 2-surface S without asking for the entire universe. The motivation is both practical and conceptual. Practically, horizons and finite extraction spheres are what numerical codes and detectors actually handle. Conceptually, a surface encloses information about curvature and trapping that cannot be reduced to a point. The difficulty is that there is no single obvious prescription that is simultaneously gauge invariant, positive, monotone under natural flows, and equal to the correct limits in every regime [9]. Different definitions trade one virtue for another.

A useful mental model is to treat quasi-local masses as different coordinate charts on the space of "energy notions". They overlap in special cases and agree in limits, but they differ away from those corners. This is not unlike how different curvature scalars emphasize different aspects of tidal geometry even though they are built from the same Riemann tensor. Mass, in this quasi-local sense, is a functional of a surface and of how that surface sits in spacetime.

12.5.2 Hawking mass

For a closed spacelike 2-surface S with area $A(S)$ and future-directed outgoing and ingoing null expansions θ_+ and θ_-, Hawking introduced

$$M_{\mathrm{H}}(S) = \sqrt{\frac{A(S)}{16\pi}} \left(1 - \frac{1}{16\pi} \int_S \theta_+ \theta_- \, dA\right). \tag{12.3}$$

This definition uses only the intrinsic area element and the null extrinsic geometry of S [8, 9]. It is sensitive to trapping: if both expansions are negative, the product $\theta_+\theta_-$ is positive and tends to increase the mass. In many asymptotically flat settings, Hawking mass approaches ADM mass for large near-round surfaces and approaches Bondi mass for appropriate families of cuts approaching null infinity. It is also closely tied to geometric flows used in inequality proofs.

At the same time, Hawking mass exposes the underlying problem. For surfaces that are wildly distorted, even in flat space, the integral term can produce behavior that is not obviously physical. This is why monotonicity and positivity statements typically require restricting the class of surfaces or the flow by which surfaces evolve. The definition is simple and geometric, but the geometry of surfaces in spacetime is subtle.

12.5.3 Misner-Sharp mass in spherical symmetry

In spherical symmetry there is a quasi-local mass with unusually clean behavior. Let R be the areal radius defined by $A = 4\pi R^2$. The Misner-Sharp mass is

$$M_{\mathrm{MS}}(R) = \frac{R}{2}\left(1 - \nabla_a R\, \nabla^a R\right), \tag{12.4}$$

in geometric units [10]. It measures how the areal radius deviates from being a simple radial coordinate. In Schwarzschild, M_{MS} equals the black hole mass for every R outside the horizon. In dynamical collapse models, its radial derivative ties directly to local energy density and work terms, which is why it often feels like the closest relativistic cousin of Newtonian enclosed mass. The limitation is also clear: the areal radius and its gradient are special to spherical symmetry.

12.5.4 Brown-York energy

Brown and York derived a quasi-local energy by Hamilton-Jacobi analysis of the gravitational action with a timelike boundary. For a 2-surface S embedded in a spacelike 3-slice, the Brown-York energy can be written schematically as

$$E_{\mathrm{BY}}(S) = \frac{1}{8\pi}\int_S (k_0 - k)\, dA, \tag{12.5}$$

where k is the trace of the extrinsic curvature of S in the physical 3-geometry and k_0 is a reference term computed from an isometric embedding of S into a reference space, often Euclidean 3-space [11]. The reference term calibrates the zero of energy and is therefore both the power and the ambiguity of the construction. In practice, Brown-York energy is attractive because it arises from boundary terms, making it natural in numerical calculations and in variational principles that keep track of boundary data.

12.6 Geometric inequalities: area bounds, censorship, and rigidity

12.6.1 The Penrose inequality as an isoperimetric statement

The Penrose inequality links horizon area to total mass. In the Riemannian time-symmetric setting, one considers an asymptotically flat 3-manifold with nonnegative scalar curvature and an outermost minimal surface representing the apparent horizon. The inequality states

$$M_{\mathrm{ADM}} \geq \sqrt{\frac{A}{16\pi}}, \tag{12.6}$$

with equality for the spatial Schwarzschild slice [12, 13, 14]. The structure resembles an isoperimetric inequality: the mass controls how much minimal-surface area can be present. Physically, Penrose argued for it by combining the idea that horizon area should not decrease classically with the expectation that singularities produced by collapse are hidden behind horizons. If those expectations hold, then an initial apparent horizon of area A cannot evolve into an asymptotically flat final state with total mass smaller than $\sqrt{A/(16\pi)}$ without exposing a naked singularity.

The inequality is also rigid. Under the hypotheses, equality forces the exterior region to be isometric to Schwarzschild. This is a recurring theme in geometric analysis: inequalities are not only bounds but also characterizations of special geometries.

12.6.2 Minimal surfaces and geometric flow intuition

Why do minimal surfaces appear. In time-symmetric data $K_{ij} = 0$, apparent horizons coincide with minimal surfaces, surfaces with zero mean curvature. They are critical points of area and act as geometric barriers. Proof strategies for the Penrose inequality use geometric flows that start at the horizon and run outward to infinity, tracking how a quasi-local mass evolves. Huisken and Ilmanen used weak inverse mean curvature flow, where surfaces expand with speed inversely proportional to mean curvature, and Bray used a conformal flow of metrics [13, 14].

For physics readers, the intuition is that the flow is a controlled way of inflating the horizon outward. One watches the transition from strong-field geometry near the trapped region to weak-field geometry near infinity, and one proves that an appropriate mass functional cannot decrease during this deformation. The theorem is the formal version of the idea that gravity cannot hide too much area behind too little mass without forcing a pathology.

12.7 Data-driven illustration: quasi-local geometry from GWTC-3 O3b events

12.7.1 Dataset and derived quantities

We now ground the discussion in an openly available dataset: the GWTC-3 O3b confident-event parameter-estimation summaries distributed through the GW Open Science Center (GWOSC) and archived as a citable snapshot on Zenodo [1, 2, 3]. For each of the 35 O3b confident events we extract median source-frame component masses (m_1, m_2), median total mass $M_{\rm tot}$, median remnant mass M_f, dimensionless remnant spin a_f, redshift z, luminosity distance D_L, and effective inspiral spin $\chi_{\rm eff}$ from the combined parameter-estimation entry.

From these medians we compute several Kerr horizon quantities in geometric units. Let

$\Delta = \sqrt{1 - a_f^2}$ and $r_+ = M_f(1 + \Delta)$. The Kerr horizon area is

$$A = 8\pi M_f^2(1 + \Delta), \tag{12.7}$$

and the irreducible mass is $M_{\rm irr} = \sqrt{A/(16\pi)}$. We also compute the surface gravity and horizon angular velocity,

$$\kappa = \frac{\Delta}{2M_f(1 + \Delta)}, \qquad \Omega_H = \frac{a_f}{2M_f^2(1 + \Delta)}. \tag{12.8}$$

These are quasi-local statements because they depend on horizon geometry rather than on the full spacetime at infinity.

To connect horizon geometry to an observable time series, we compute the dominant Kerr ringdown frequency f_{220} and damping time τ_{220} using standard fits of the dimensionless quasinormal frequency and quality factor as functions of spin. The fit coefficients used here are publicly provided in a plain-text file [15] and are consistent with the spectroscopy literature [16]. Converting mass in solar masses to seconds uses $M_\odot G/c^3 \approx 4.925\,490\,95 \times 10^{-6}\,\mathrm{s}$.

Because we compute derived quantities from reported medians, some derived relationships that should hold exactly at the level of full posteriors become approximate in this deterministic reconstruction. This is expected and not a flaw: it is precisely the kind of mathematical modification that turns a catalog into raw material for geometric illustration. For high-precision statements one should propagate full posterior samples, but for qualitative geometric structure, the median-based reconstruction is sufficient and reproducible.

12.7.2 Ten data-based figures

Each of the following ten figures is generated directly from the embedded dataset and analytic transformations such as Eqs. (12.7) and (12.8).

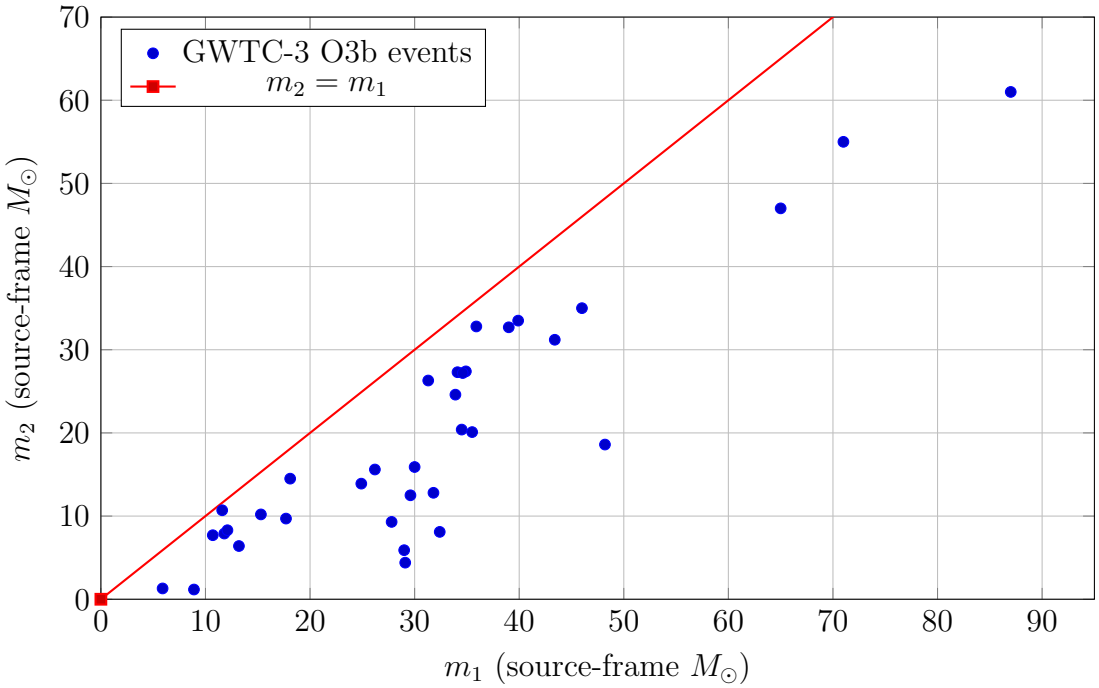


Figure 12.1: Component-mass scatter for GWTC-3 O3b confident events (median source-frame masses), showing the range of mass ratios present in the public sample [1, 2].

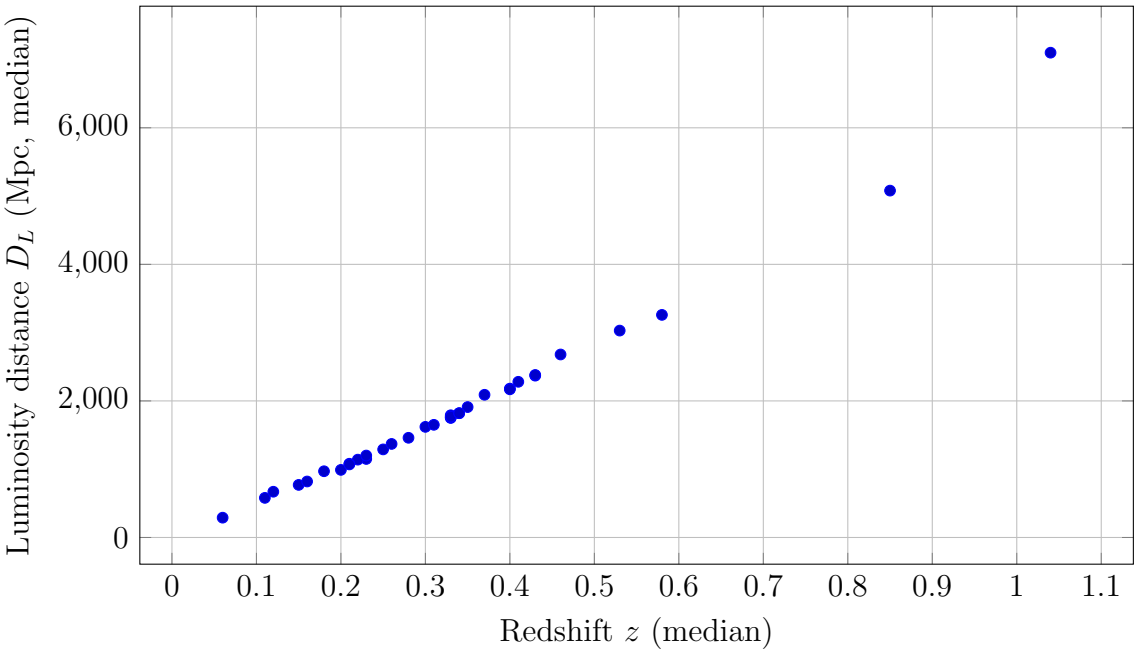


Figure 12.2: Distance scale of the same event set: median luminosity distance versus redshift as reported by GWOSC [1].

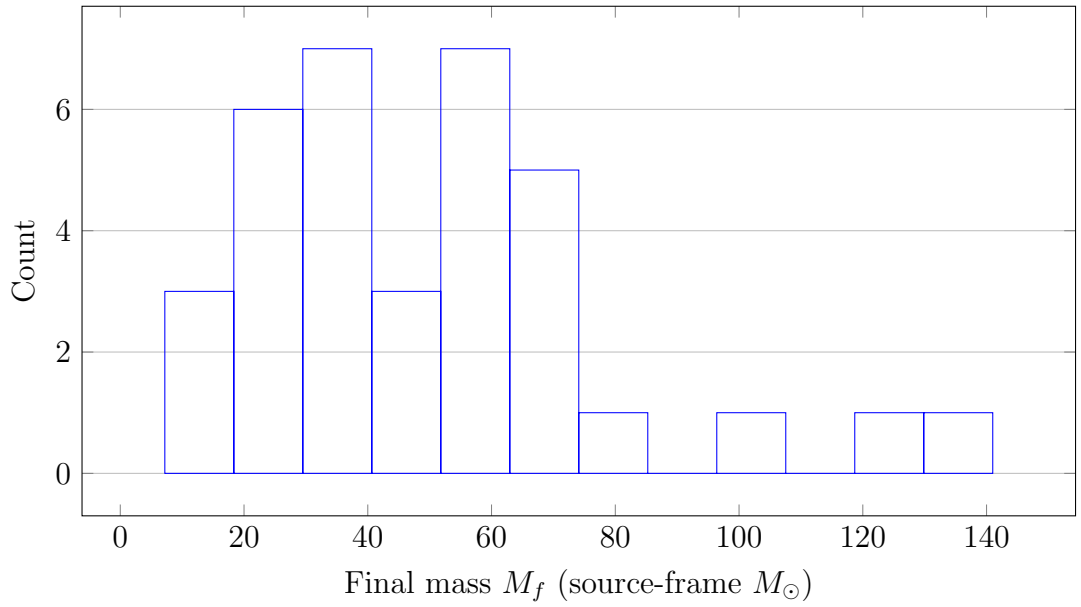


Figure 12.3: Histogram of remnant masses M_f for the GWTC-3 O3b confident sample [1, 2].

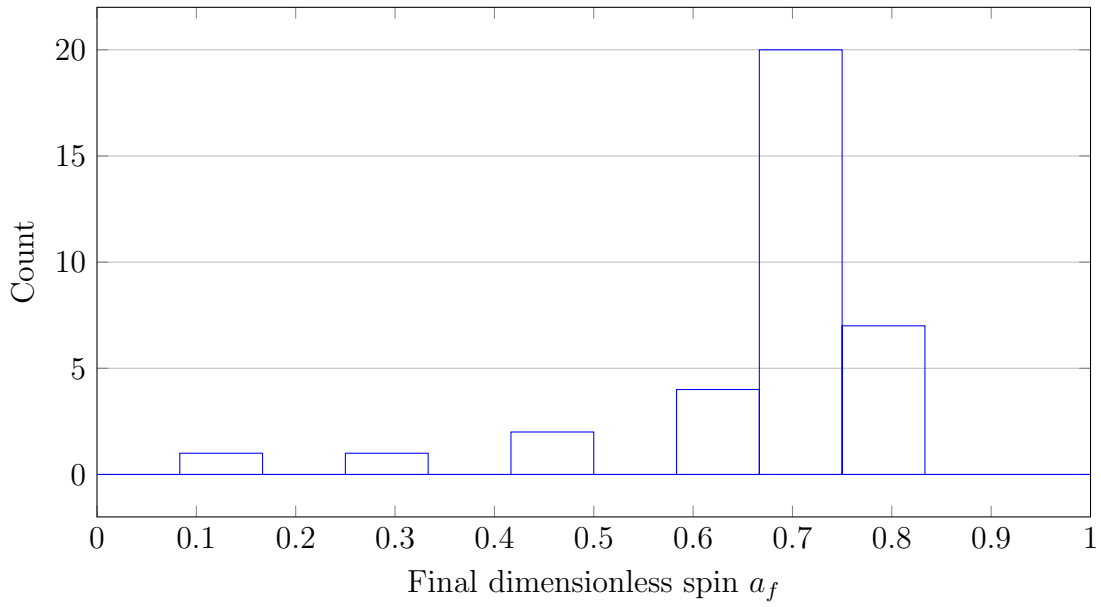


Figure 12.4: Histogram of remnant spins a_f. Spin is the control parameter for Kerr horizon area, surface gravity, and quasinormal spectra [1].

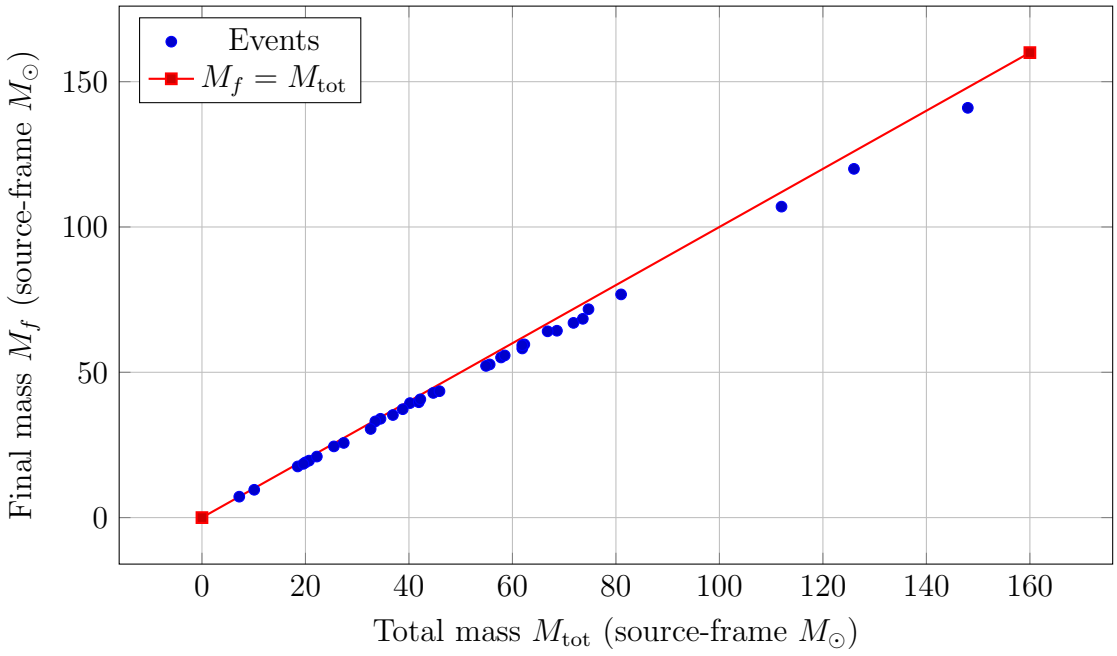


Figure 12.5: Final mass compared with total source-frame mass. The gap below the diagonal represents mass-energy radiated in gravitational waves, estimated from medians [1].

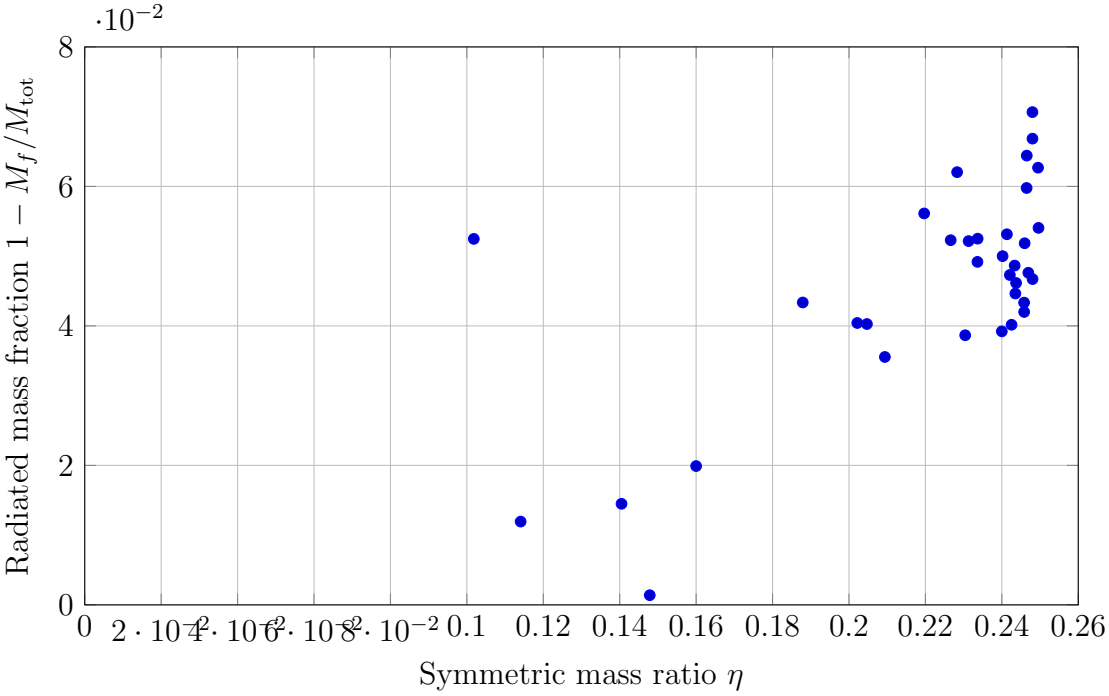


Figure 12.6: Estimated radiated mass fraction versus symmetric mass ratio $\eta = (m_1 m_2)/(m_1 + m_2)^2$. The fraction is computed deterministically from reported medians [1].

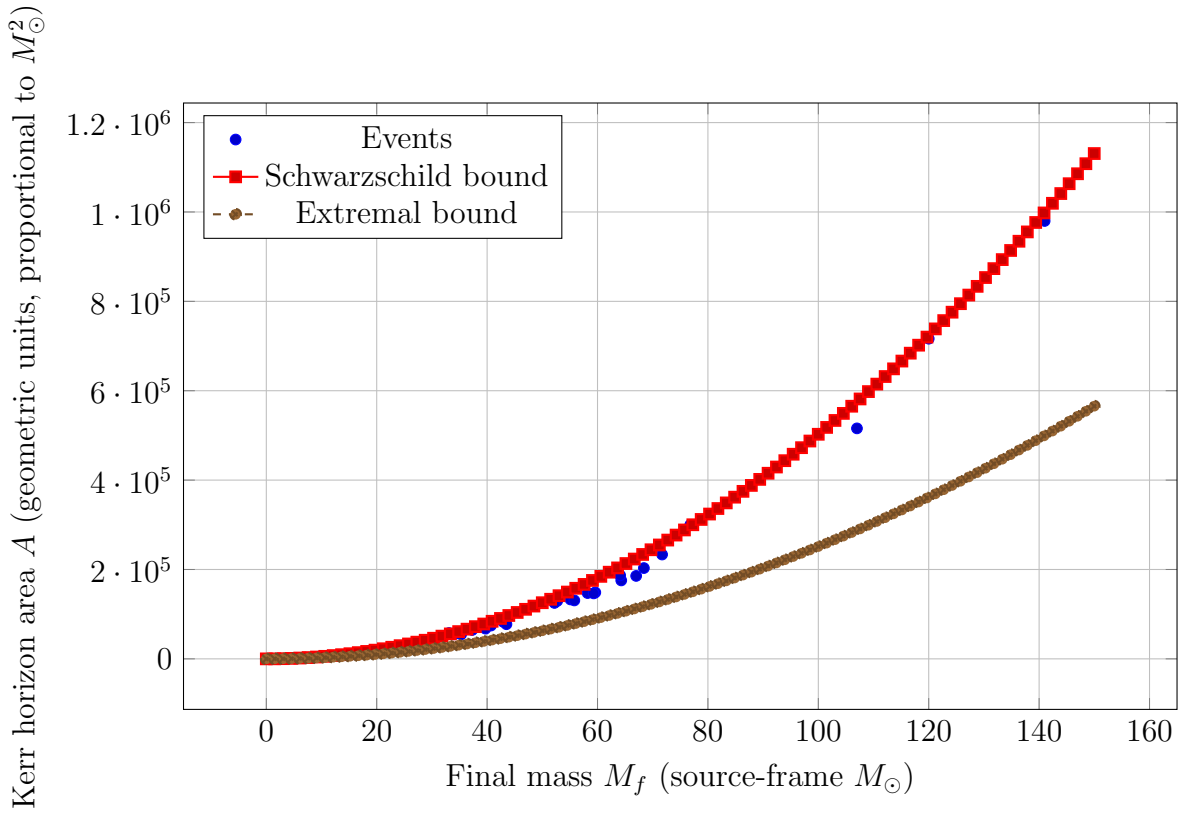


Figure 12.7: Horizon area computed from remnant (M_f, a_f) using Eq. (12.7). The curves show the Schwarzschild upper bound ($a_f = 0$) and the extremal lower bound (near $a_f = 1$) for fixed mass.

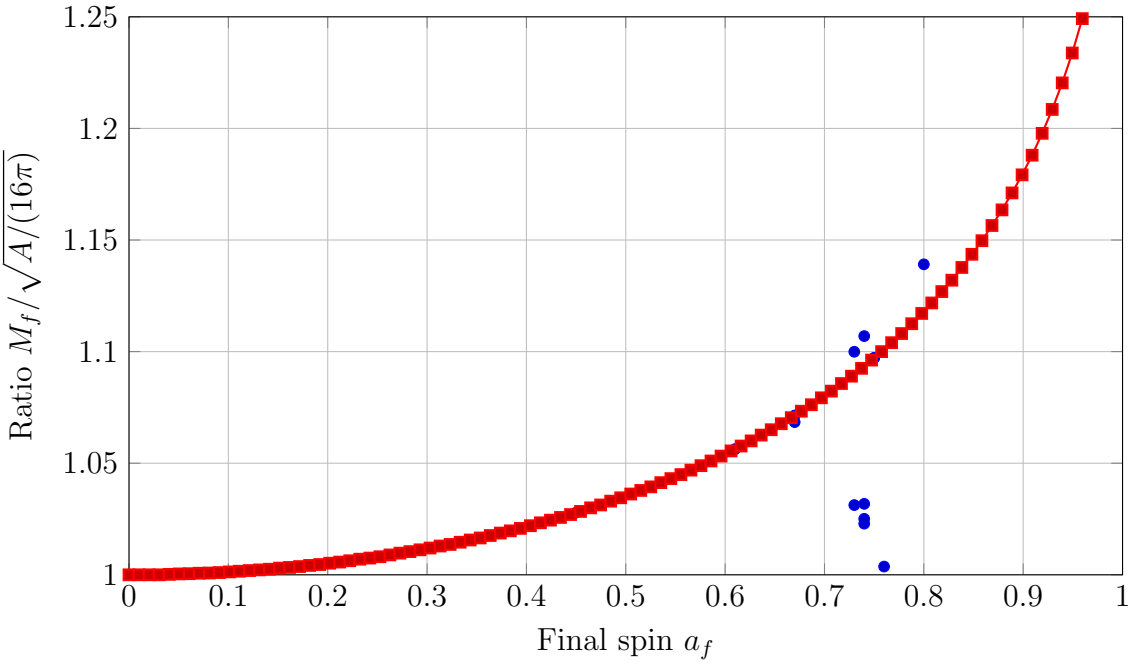


Figure 12.8: A Kerr-specific Penrose-type ratio: M_f/M_{irr} where $M_{\text{irr}} = \sqrt{A/(16\pi)}$. The analytic curve shows the exact Kerr relation as a function of spin; points are computed from event medians.

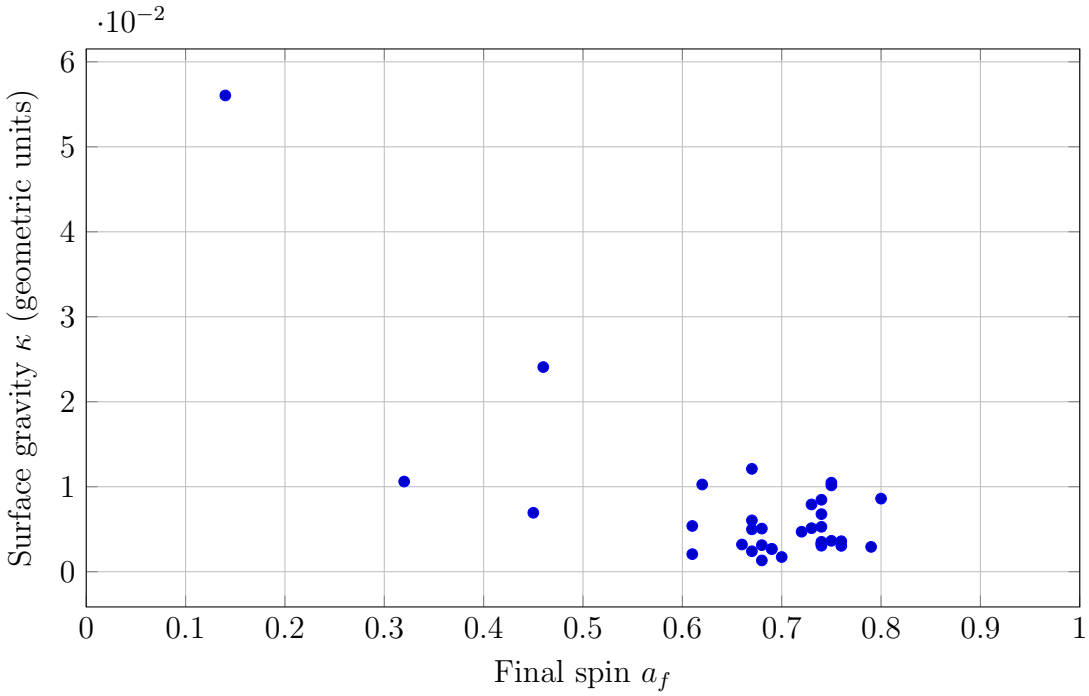


Figure 12.9: Surface gravity κ computed from the Kerr formula in Eq. (12.8) using event medians. The decreasing trend with increasing a_f illustrates the approach toward the extremal limit where $\kappa \to 0$.

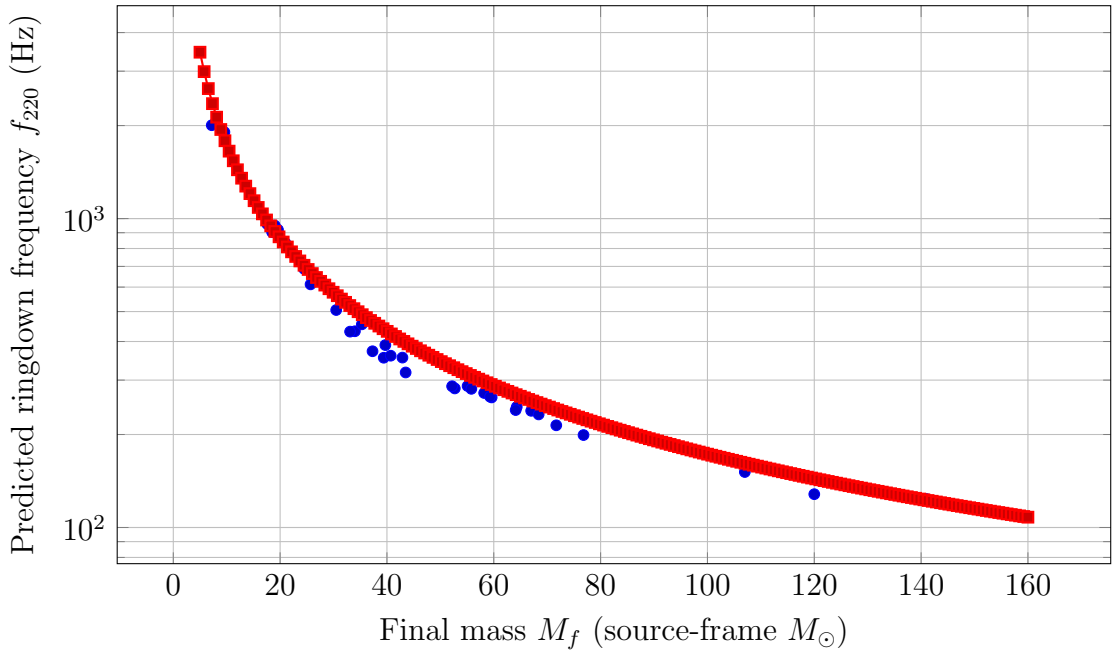


Figure 12.10: Predicted dominant Kerr ringdown frequency f_{220} from remnant (M_f, a_f) using public fit coefficients [15]. Points are event-based values; the curve shows the scaling for a representative fixed spin $a_f = 0.7$ to emphasize the approximate $1/M$ behavior [16].

12.8 Interpretation: what the figures are telling us

The catalog scatter in Fig. 12.1 is more than event bookkeeping. It is a geometric input table. Once (m_1, m_2) are known, invariants like η and the chirp mass follow, and these invariants control the inspiral phasing. But the same numbers also help determine the remnant state. A large mass ratio, visible as points far below the diagonal, tends to produce lower remnant spin than equal-mass mergers, and therefore tends to produce larger surface gravity at fixed final mass. This is one way the chapter's geometric emphasis connects directly to waveform morphology.

Figure 12.2 reminds us that quasi-local geometry and asymptotic structure coexist in data products. The horizon quantities computed later depend on the source-frame mass scale, which depends on cosmological redshift. A 10 percent shift in inferred redshift shifts the inferred source-frame mass, and therefore shifts A, κ, and f_{220} coherently. In that sense, the chapter's theme that mass depends on how you ask the question is not philosophical. It is a practical statement about parameter inference.

The mass and spin histograms in Figs. 12.3 and 12.4 show that O3b remnants mostly live in a moderate-spin regime rather than near extremality. In inequality language, that means these events typically do not probe the most delicate saturation regimes of bounds. However, they do probe the basic wedge between Schwarzschild and extremal Kerr, which is enough to show how spin deforms area at fixed mass.

Figures 12.5 and 12.6 translate radiation into a geometric accounting identity. Even when computed from medians, the radiated fraction is a few percent for many events. This is observationally consistent with the idea that a merger sheds a modest fraction of the total mass-energy into gravitational waves, while the horizon area grows. Because area scales quadratically with mass, modest differences in mass can correspond to large differences in area. That amplification is why inequalities that involve area are so effective.

Figure 12.7 is an inequality picture in disguise. For fixed M_f, the Kerr area lies between $8\pi M_f^2$ and $16\pi M_f^2$. In the language of geometric analysis, rotation provides a deformation direction that reduces the area. In the language of black hole mechanics, it reduces the irreducible mass for fixed total mass. Figure 12.8 then packages the same information in a Penrose-like ratio $M_f/M_{\rm irr}$. The curve is not fitted to the data; it is an exact Kerr identity. The point cloud follows it because the points are constructed from the same identity using measured (M_f, a_f). This is still useful because it turns catalog numbers into a single geometric diagnostic.

Finally, Fig. 12.10 connects quasi-local horizon geometry to an observable frequency scale. The ringdown frequency is essentially set by $1/M_f$ with a spin-dependent correction. That is why the plot is close to a power law in M_f and why the y-axis is shown in log scale. This is the simplest example of a geometric pipeline: start from a quasi-local horizon state, output a predicted spectral line.

12.9 Real-world applicability: phenomenon statements

12.9.1 Phenomenon 1: numerical relativity sanity checks as inequality checks

In numerical relativity, one evolves discretized initial data and repeatedly locates apparent horizons as marginally trapped surfaces. Every time a horizon is found, one can compute its area and a quasi-local spin estimate, yielding an irreducible mass and related horizon invariants. Geometric inequalities then become unit tests for the simulation. If a run violates a known inequality under conditions where it should hold, the violation is almost never new physics. It is a diagnostic of gauge pathology, insufficient resolution, or an error in the horizon finder. The reason this works is that inequalities are global constraints: they are hard to satisfy accidentally if the geometry is numerically wrong, and easy to satisfy once the geometry is correct. This makes inequalities a powerful correctness criterion, independent of the detailed waveform.

12.9.2 Phenomenon 2: gravitational-wave catalogs as horizon-geometry catalogs

A gravitational-wave catalog is often described as a list of masses and spins, but this chapter suggests a stronger interpretation: it is a list of horizon geometries. Each remnant mass and spin corresponds to a Kerr horizon with definite area, surface gravity, and angular velocity. Those numbers are the state variables that appear in the first law of black hole mechanics and in thermodynamic analogies. Therefore the catalog can be treated as a thermodynamic dataset: one can ask how area and surface gravity are distributed, how they correlate with inferred mass ratio, and how close or far the population sits from extremality. This reframing is operational, not metaphorical, because these quantities are computed directly from measured remnant parameters.

12.9.3 Phenomenon 3: multi-messenger calibration and the meaning of mass at infinity

When an event has host-galaxy identification or an electromagnetic counterpart, one can obtain an independent redshift. That redshift recalibrates detector-frame masses into source-frame masses and therefore shifts the inferred horizon area and surface gravity. In other words, cosmological information influences quasi-local geometry through the mass scale. This provides a concrete example of how asymptotic and quasi-local notions are intertwined in practice. The horizon is a quasi-local object, but the number we assign to its mass is informed by how the spacetime is embedded in a cosmological setting and by how the signal reaches future null infinity.

12.9.4 Phenomenon 4: horizon growth and energy flux as a two-sided ledger

In radiating spacetimes, the Bondi mass decreases at null infinity while horizon area increases under classical energy conditions. That is a two-sided energy ledger: energy leaves the system through null infinity and the horizon responds by growing. Modern observations allow partial access to both sides. Inspiral and merger encode radiated energy in the waveform amplitude and phase evolution, while remnant parameters encode the final horizon area. Even when one does not reconstruct the full Bondi news tensor, the concept explains why quasi-local and asymptotic masses are complementary rather than redundant. One describes what is left inside, the other describes what escaped.

12.9.5 Phenomenon 5: constraints on exotic alternatives to black holes

Any exotic compact object proposed as an alternative to a classical black hole must reconcile the same geometric accounting. It must reproduce mass notions at infinity, it must admit sensible quasi-local energies for finite surfaces, and it must explain the observed ringdown behavior. If it lacks an event horizon, it must explain why trapped surfaces are absent or modified, yet it must also explain why inferred remnants behave like Kerr in the regime where detectors are sensitive. Inequalities sharpen this challenge because they tie area-like quantities to mass-like quantities in a way that is difficult to mimic without an actual horizon. In this sense, geometric inequalities are not only mathematical results. They are discriminants for model-building.

12.10 Conceptual diagrams (non-dataset figures)

The following figures are schematic conceptual aids. They are included to clarify definitions and proof ideas. They are not counted among the chapter's ten data-driven illustrations.

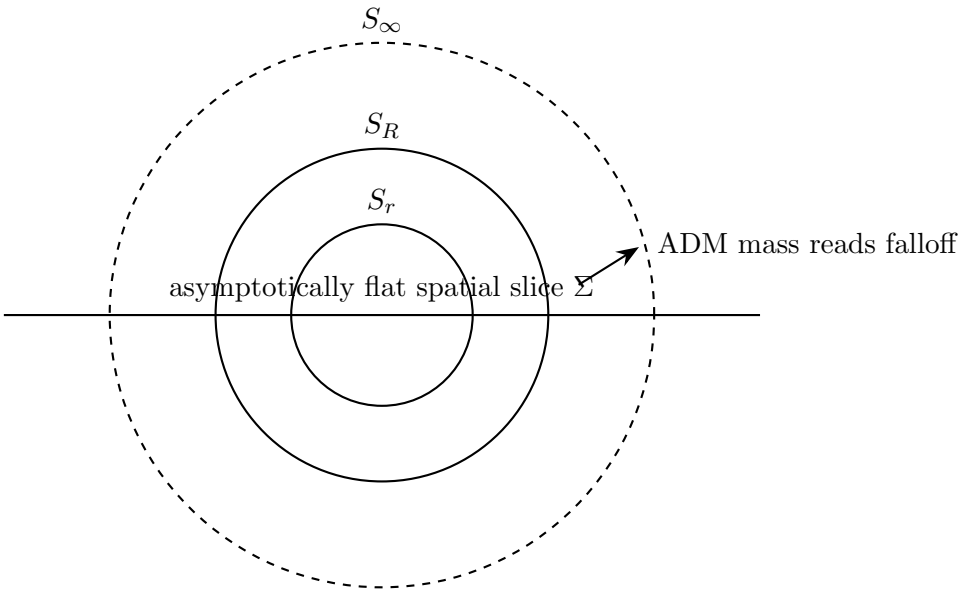


Figure 12.11: Conceptual diagram: ADM mass as a surface integral at spatial infinity. Compare with Eq. (12.1).

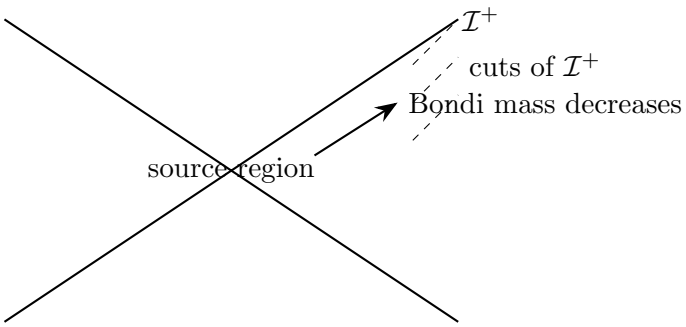


Figure 12.12: Conceptual diagram: Bondi mass is defined at future null infinity and decreases when gravitational radiation is present [5].

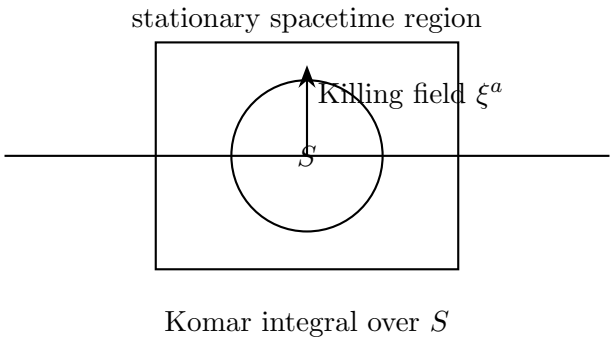


Figure 12.13: Conceptual diagram: Komar mass from a Killing field and a 2-surface integral, Eq. (12.2) [7].

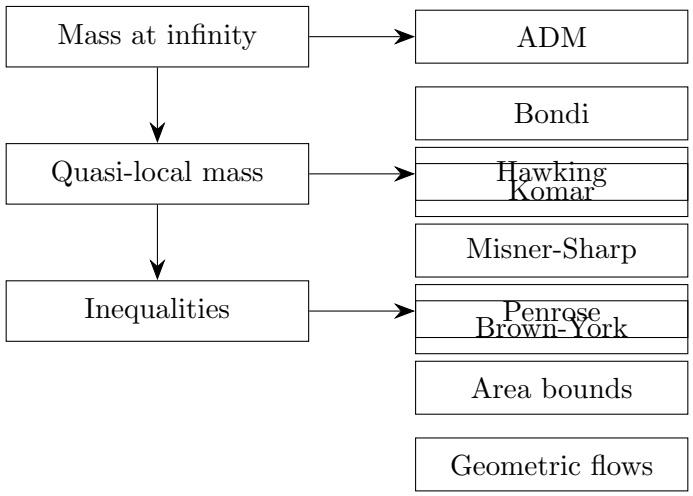


Figure 12.14: Conceptual diagram: a map of definitions and where they live [9, 12].

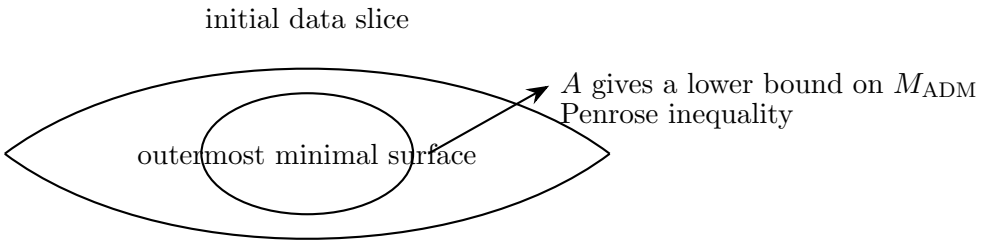


Figure 12.15: Conceptual diagram: Penrose inequality as a geometric isoperimetric constraint [13, 14].

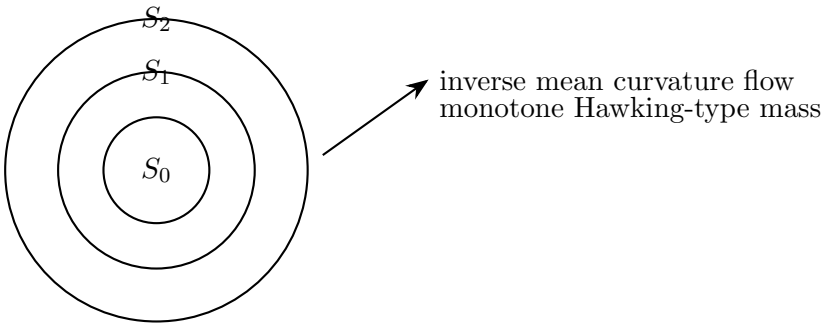


Figure 12.16: Conceptual diagram: inverse mean curvature flow expands surfaces outward while controlling a Hawking-type mass, a key idea in proofs of the Riemannian Penrose inequality [13].

12.11 Conclusion

Quasi-local mass is the geometric answer to a physical question that cannot be posed locally in general relativity. ADM and Bondi mass show how total energy can be measured at infinity, with Bondi mass explicitly tracking radiative loss in retarded time. Komar integrals show that in stationary settings, mass is a symmetry charge associated with a timelike Killing field. Quasi-local candidates such as Hawking mass, Misner-Sharp mass, and Brown-York energy then push energy accounting onto finite surfaces, which is where horizons live and where numerical and observational workflows operate.

Geometric inequalities, especially the Penrose inequality, convert these mass notions into statements that constrain geometry and encode deep expectations about gravitational collapse and censorship. The data-based illustrations demonstrate that the chapter is not purely theoretical: public gravitational-wave catalogs already contain horizon geometry in compressed form. From observed remnant mass and spin one can compute area, irreducible mass, surface gravity, and ringdown spectra, turning quasi-local geometry into a reproducible pipeline from data to curvature-informed interpretation.

Appendix A

Dataset table (embedded in this manuscript)

A.1 GWTC-3 O3b confident-event medians used for plots

Table A.1 lists the primary catalog columns embedded in this manuscript and used to generate Figs. 12.1 to 12.10. Values are medians as reported by GWOSC for the combined parameter-estimation entry for each event in GWTC-3 O3b [1]. Derived columns used in plots are deterministic functions of these medians and of the Kerr and QNM fit formulas.

#	Event	m_1	m_2	M_f	a_f	z	D_L (Mpc)
1	GW191103_012549	11.8	7.9	19	0.75	0.2	990
2	GW191105_143521	10.7	7.7	17.6	0.67	0.23	1150
3	GW191109_010717	65	47	107	0.61	0.25	1290
4	GW191113_071753	29	5.9	34	0.45	0.26	1370
5	GW191126_115259	12.1	8.3	19.6	0.75	0.3	1620
6	GW191127_050227	27.8	9.3	35.3	0.67	0.46	2680
7	GW191129_134029	11.6	10.7	21	0.8	0.15	770
8	GW191204_110529	17.7	9.7	25.7	0.73	0.33	1790
9	GW191215_223052	32.4	8.1	39.4	0.61	0.37	2090
10	GW191216_213338	13.2	6.4	18.5	0.62	0.11	580
11	GW191219_163120	31.8	12.8	42.9	0.67	0.4	2180
12	GW191222_033537	46	35	76.8	0.67	0.33	1750
13	GW200105_162426	8.9	1.17	9.57	0.46	0.12	670
14	GW200112_155838	29.1	4.4	33.1	0.32	0.33	1760
15	GW200115_042309	5.9	1.3	7.2	0.14	0.06	290
16	GW200128_022011	30	15.9	43.5	0.72	0.22	1140
17	GW200129_065458	39	32.7	67	0.79	0.21	1080
18	GW200202_154313	39.9	33.5	68.4	0.76	0.28	1460
19	GW200208_130117	34.1	27.3	58.2	0.76	0.35	1910
20	GW200209_085452	43.4	31.2	71.7	0.69	0.43	2370
21	GW200210_092254	24.9	13.9	37.3	0.74	0.18	970
22	GW200216_220804	71	55	120	0.7	0.58	3260
23	GW200220_061928	26.2	15.6	39.7	0.73	0.33	1770
24	GW200224_222234	35.9	32.8	64.3	0.74	0.34	1820
25	GW200225_060421	18.1	14.5	30.5	0.74	0.31	1650
26	GW200302_015811	35.5	20.1	52.7	0.66	0.43	2380
27	GW200306_093714	34.5	20.4	52.2	0.68	0.41	2280
28	GW200308_173609	48.2	18.6	64.1	0.69	0.85	5080
29	GW200311_115853	34.9	27.4	59.6	0.74	0.53	3030
30	GW200316_215756	15.3	10.2	24.5	0.74	0.16	820
31	GW200322_091133	34.6	27.2	59.3	0.74	0.4	2170
32	GW200323_004856	87	61	141	0.68	1.04	7100
33	GW200329_093316	31.3	26.3	55.1	0.75	0.23	1200
34	GW200330_063833	29.6	12.5	40.7	0.68	0.21	1070
35	GW200403_194701	33.9	24.6	55.8	0.74	0.34	1820

Table A.1: Subset of embedded GWTC-3 O3b medians from the GWOSC snapshot dataset [1].

Bibliography

[1] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, "GWOSC Event Portal Snapshots," Zenodo, ver. v17, Aug. 2025. DOI: 10.5281/zenodo.10091305. https://zenodo.org/records/10091305

[2] R. Abbott *et al.*, "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run," *Phys. Rev. X*, vol. 13, 041039, 2023. DOI: 10.1103/PhysRevX.13.041039. https://arxiv.org/abs/2111.03606

[3] GW Open Science Center, "Event Portal Usage Notes," accessed Feb. 24, 2026. https://gwosc.org/eventportaldocs/

[4] R. Arnowitt, S. Deser, and C. W. Misner, "The Dynamics of General Relativity," arXiv:gr-qc/0405109, 2004. https://arxiv.org/abs/gr-qc/0405109

[5] J. Winicour, "Bondi-Sachs Formalism," arXiv:1609.01731, 2016. https://arxiv.org/pdf/1609.01731

[6] H. Bondi, "Gravitational Waves in General Relativity," *Nature*, vol. 186, p. 535, 1960. https://www.nature.com/articles/186535a0.pdf

[7] A. Komar, "Positive-Definite Energy Density and Global Consequences for General Relativity," *Phys. Rev.*, vol. 129, pp. 1873-1876, 1963. DOI: 10.1103/PhysRev.129.1873. https://harvest.aps.org/v2/journals/articles/10.1103/PhysRev.129.1873/fulltext

[8] S. W. Hawking, "Gravitational Radiation in an Expanding Universe," *J. Math. Phys.*, vol. 9, no. 4, pp. 598-604, 1968. DOI: 10.1063/1.1664615. https://pubs.aip.org/aip/jmp/article/9/4/598/234155/Gravitational-Radiation-in-an-Expanding-Universe

[9] L. B. Szabados, "Quasi-Local Energy-Momentum and Angular Momentum in General Relativity: A Review Article," *Living Rev. Relativity*, vol. 7, no. 4, 2004. DOI: 10.12942/lrr-2004-4. https://link.springer.com/article/10.12942/lrr-2004-4

[10] C. W. Misner and D. H. Sharp, "Relativistic Equations for Adiabatic, Spherically Symmetric Gravitational Collapse," *Phys. Rev.*, vol. 136, pp. B571-B576, 1964. DOI: 10.1103/PhysRev.136.B571. https://harvest.aps.org/v2/journals/articles/10.1103/PhysRev.136.B571/fulltext

[11] J. D. Brown and J. W. York, "Quasilocal energy and conserved charges derived from the gravitational action," *Phys. Rev. D*, vol. 47, pp. 1407-1419, 1993. DOI: 10.1103/PhysRevD.47.1407. https://link.aps.org/doi/10.1103/PhysRevD.47.1407

[12] H. Bray and P. Chruściel, "The Penrose Inequality," arXiv:gr-qc/0312047, 2003. https://arxiv.org/abs/gr-qc/0312047

[13] G. Huisken and T. Ilmanen, "The inverse mean curvature flow and the Riemannian Penrose Inequality," 2001. https://math.jhu.edu/~js/Math646/huisken-ilmanen.pdf

[14] H. L. Bray, "Proof of the Riemannian Penrose inequality using the Positive Mass Theorem," *J. Differential Geom.*, vol. 59, pp. 177-267, 2001. https://math.jhu.edu/~js/Math646/bray.penrose.pdf

[15] E. Berti, "Fits to Kerr QNM frequencies (plain-text coefficients)," accessed Feb. 24, 2026. https://pages.jh.edu/~eberti2/ringdown/fitcoeffsWEB.dat

[16] E. Berti, V. Cardoso, and C. M. Will, "On Gravitational-Wave Spectroscopy of Massive Black Holes with the Space Interferometer LISA," *Phys. Rev. D*, vol. 73, 064030, 2006. DOI: 10.1103/PhysRevD.73.064030. https://link.aps.org/doi/10.1103/PhysRevD.73.064030

[17] P. T. Chruściel, "Mass in General Relativity," in *The Global Nonlinear Stability of the Kerr Black Hole*, Springer, 2021, Ch. 6 (PDF). https://link.springer.com/content/pdf/10.1007/978-3-030-65683-6_6.pdf

Black Holes Geometry: Curvature and Horizons

Chapter 13: 3+1 Geometry and the Initial Value Problem

Abstract. The Einstein equations describe spacetime as geometry, but the geometry is dynamical: curvature evolves, horizons form, and gravitational radiation carries information to infinity. The 3+1 decomposition is the geometric device that turns "spacetime as a whole" into an initial value problem: provide a three dimensional slice with suitable geometric data, then evolve it consistently in time. This chapter builds a coordinate aware yet coordinate respectful account of foliations, lapse and shift, induced metrics, and extrinsic curvature. It then connects Gauss Codazzi relations to the Hamiltonian and momentum constraints, explains how constraint solving becomes an elliptic geometry problem, and interprets evolution equations as gauge driven rules for how a slicing moves through the four dimensional manifold. Apparent horizon finding is presented as a quasi local geometric partial differential equation, highlighting why horizons are natural objects in 3+1 language. Finally, the chapter uses an open real world dataset of gravitational wave events from the GWTC 3 catalog to create data based illustrations that tie abstract variables like masses, effective spin, distance, and remnant mass to the geometric story of initial data and horizon formation. The figures are based on catalog data and on derived quantities computed directly from those data.

Keywords: 3+1 decomposition, ADM formalism, lapse, shift, extrinsic curvature, Gauss Codazzi, constraint equations, initial data, numerical relativity, apparent horizon finding, GWTC 3, GWOSC event portal

Contents

13 3+1 Geometry and the Initial Value Problem **1**

13.1 Introduction: why split spacetime at all? 1

13.2 Foliations, lapse, shift, and the induced three metric 2

13.2.1 Conceptual diagram: slicing, lapse, and shift 3

13.2.2 Extrinsic curvature as the time derivative of geometry 3

13.3 Gauss Codazzi relations and the constraint equations 4

13.3.1 Conceptual diagram: projection geometry behind the constraints . 5

13.3.2 Why constraints are geometry, not bookkeeping 5

13.4 Solving the constraints: conformal geometry as an elliptic engine 6

13.4.1 Conceptual diagram: constraint solving pipeline 7

13.5 Evolution equations and the meaning of gauge choices 7

13.5.1 Conceptual diagram: lapse and shift as a controlled deformation . 8

13.5.2 Conceptual diagram: evolution, gauge, and constraint monitoring loop 9

13.6 Apparent horizons as quasi local geometry on a slice 9

13.6.1 Conceptual diagram: apparent horizon finding on a slice 10

13.7 Data based illustrations: catalog level shadows of 3+1 geometry 10

13.7.1 Dataset provenance and parameter meanings 11

13.7.2 Interpreting the dataset through a 3+1 lens 12

13.8 Practical geometric pitfalls: what can go wrong, and why it is geometric 17

13.9 Conclusion 18

Chapter 13

3+1 Geometry and the Initial Value Problem

13.1 Introduction: why split spacetime at all?

General relativity is often introduced as a four dimensional story: spacetime is a Lorentzian manifold and matter tells curvature how to bend. That point of view is conceptually clean, but it hides a practical question that becomes unavoidable the moment we ask for dynamics. If curvature can change, then what does it mean to specify "the gravitational field now" and predict it "later"? The 3+1 decomposition answers this by insisting on a geometric act of bookkeeping: choose a foliation of spacetime by spacelike hypersurfaces and describe the geometry of each hypersurface together with how it sits inside spacetime. The result is a precise initial value formulation, in which initial data are not arbitrary, but must satisfy constraint equations that are themselves geometric consequences of the Einstein equations.

This chapter is written with a deliberate tension in mind. On the one hand, the foliation introduces coordinates and slicing choices that can feel like "mere gauge." On the other hand, the foliation is not a computational trick; it is a genuine geometric structure that lets us express the dynamics in terms of intrinsic three geometry and extrinsic embedding geometry. Once you develop the right intuition, lapse and shift are not annoyances but levers: they control how the slicing threads through spacetime and how the coordinate grid moves relative to the physical geometry.

The most concrete reason 3+1 geometry matters is that it underlies essentially all modern time domain modeling of black hole dynamics, from numerical simulations of mergers to apparent horizon tracking and quasi local diagnostics. Yet, the purpose of this chapter is not to turn the reader into a numerical relativist overnight. Instead, the goal is to

make the geometric meaning of the initial value problem feel inevitable: constraints as compatibility conditions, evolution as controlled deformation, and horizons as quasi local objects naturally expressed on a slice.

Phenomenon statement 1 (observational relevance). Gravitational wave detectors measure time series data on Earth, but the theoretical prediction for those time series is generated by evolving initial data for a binary system through a strongly curved merger. The mapping from "two compact objects at some separation" to "a waveform at infinity" is a 3+1 pipeline: constraint solving produces consistent initial data, evolution equations propagate the geometry forward, and quasi local horizon finding provides remnant properties such as final mass. Catalog level event parameters are therefore a data level shadow of a geometric initial value problem.

13.2 Foliations, lapse, shift, and the induced three metric

Let $(\mathcal{M}, g_{ab})$ be a four dimensional spacetime with Lorentzian metric signature $(-,+,+,+)$. A 3+1 decomposition begins with a foliation by spacelike hypersurfaces Σ_t, labeled by a scalar function $t : \mathcal{M} \to \mathbb{R}$ whose level sets are spacelike. The unit timelike normal to the hypersurfaces is n^a, satisfying

$$g_{ab} n^a n^b = -1, \qquad n_a = -\alpha \nabla_a t, \tag{13.1}$$

where α is the *lapse function*. The lapse is a scalar field that converts an increment in coordinate time into the proper time separation between neighboring hypersurfaces along the normal direction. In this sense, α measures how quickly the foliation advances in the normal direction.

The induced three metric γ_{ab} on each slice is the projection of the spacetime metric orthogonal to n^a:

$$\gamma_{ab} = g_{ab} + n_a n_b. \tag{13.2}$$

This tensor is purely spatial in the sense that $\gamma_{ab} n^b = 0$. It is the geometric object that measures distances and angles *within* a slice, and it is the primary unknown in the initial value problem.

The second core ingredient is the *shift vector* β^a, which lives tangent to the slices. The coordinate time flow vector field t^a that advances points from one slice to the next

decomposes as

$$t^a = \alpha n^a + \beta^a, \qquad \beta^a n_a = 0. \tag{13.3}$$

Geometrically, αn^a tells you how much you move orthogonally off the slice when increasing t, while β^a tells you how the spatial coordinates slide within the slice during that same advance. This split is not an arbitrary trick: it is the statement that "time evolution" in coordinates can be decomposed into a normal deformation and a tangential relabeling.

13.2.1 Conceptual diagram: slicing, lapse, and shift

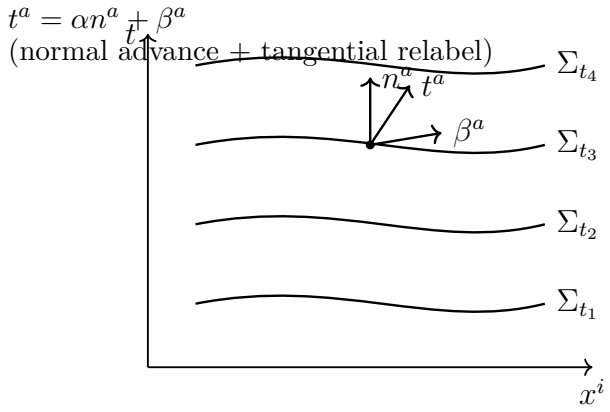


Figure 13.1: **Conceptual diagram (non data).** A foliation of spacetime by spacelike hypersurfaces Σ_t. The coordinate time flow t^a decomposes into a normal part controlled by the lapse α and a tangential part given by the shift β^a. This is the geometric basis of the 3+1 split [1, 2].

13.2.2 Extrinsic curvature as the time derivative of geometry

If γ_{ab} is the intrinsic metric on a slice, then the *extrinsic curvature* K_{ab} measures how the slice bends inside the ambient spacetime. One clean definition is

$$K_{ab} := -\gamma_a{}^c \gamma_b{}^d \nabla_c n_d. \tag{13.4}$$

This is symmetric when the foliation is hypersurface orthogonal, which we assume here. A second definition, crucial for intuition, is that K_{ab} is essentially the Lie derivative of γ_{ab} along the normal:

$$K_{ab} = -\frac{1}{2}\mathcal{L}_n \gamma_{ab}. \tag{13.5}$$

This makes K_{ab} feel like a geometric "velocity" of the three metric: it encodes how the spatial geometry changes from slice to slice when you move orthogonally through spacetime. In a Hamiltonian language, (γ_{ij}, K_{ij}) plays the role of position and momentum.

Phenomenon statement 2 (what "time dilation" means in 3+1 language). In a black hole spacetime, different choices of lapse correspond to different clocks that label the same physical geometry. A small lapse near a horizon means adjacent coordinate time slices are separated by little proper time along the normal. This is not merely a computational quirk: it encodes how a chosen foliation avoids or approaches strong curvature regions, affecting what "now" means near the horizon.

13.3 Gauss Codazzi relations and the constraint equations

The Einstein equations are four dimensional tensor equations. When we split spacetime into space plus time, we must ask what parts of those equations are evolution equations and what parts are constraints on initial data. The answer is geometric: it comes from projecting curvature identities onto and along the hypersurfaces. The Gauss and Codazzi relations relate the four dimensional Riemann tensor to the intrinsic three dimensional curvature of Σ_t and the extrinsic curvature.

Let D_i be the covariant derivative compatible with γ_{ij}, and let ${}^{(3)}R$ be the scalar curvature of the slice. In vacuum or with matter, define the energy density and momentum density measured by observers with four velocity n^a:

$$\rho := T_{ab} n^a n^b, \qquad S_i := -T_{ab} n^a \gamma^a{}_i. \tag{13.6}$$

Then the Einstein equations projected twice along n^a and twice onto the slice yield the *Hamiltonian constraint*:

$${}^{(3)}R + K^2 - K_{ij}K^{ij} = 16\pi\rho, \tag{13.7}$$

where $K := \gamma^{ij}K_{ij}$ is the trace. Projecting once along n^a and once onto the slice yields the *momentum constraint*:

$$D_j \left(K^{ij} - \gamma^{ij} K \right) = 8\pi S^i. \tag{13.8}$$

These constraints are not optional. If they are violated on the initial slice, the evolved spacetime will not satisfy the Einstein equations, no matter how carefully you integrate the evolution system.

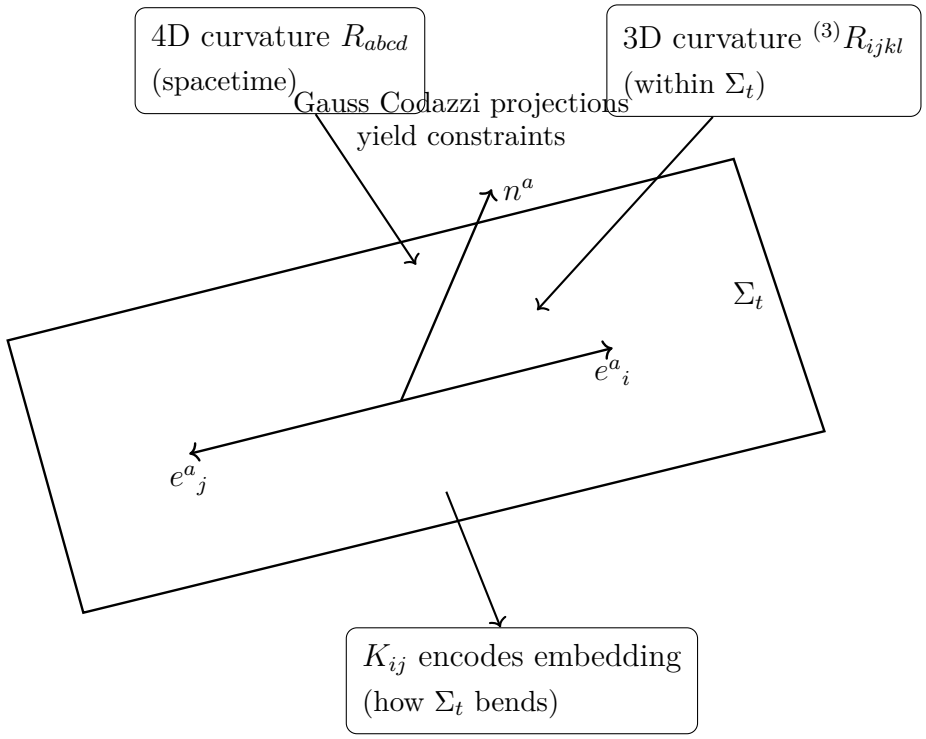


Figure 13.2: **Conceptual diagram (non data).** The constraint equations arise from projecting spacetime curvature onto the hypersurface Σ_t and along its unit normal n^a. Intrinsic curvature ${}^{(3)}R_{ijkl}$ and extrinsic curvature K_{ij} together encode the relevant components of spacetime curvature in 3+1 form [1, 2].

13.3.1 Conceptual diagram: projection geometry behind the constraints

13.3.2 Why constraints are geometry, not bookkeeping

Equations (13.7) and (13.8) are sometimes described as "constraints" because they do not involve second time derivatives. A better geometric way to say it is: they are compatibility conditions between intrinsic and extrinsic geometry. If you specify γ_{ij} and K_{ij} arbitrarily, there is no reason they should come from embedding a three manifold as a slice of a four manifold satisfying Einstein's equations. The constraints are the obstructions to that embedding.

The constraints are also where black hole geometry first becomes tangible in the initial value formulation. A trapped surface, for example, is a two dimensional surface inside Σ_t whose outward and inward null expansions are both negative. Whether such a surface exists depends on the interplay between the intrinsic curvature of the slice and the extrinsic curvature, because expansions involve both γ_{ij} and K_{ij}. That is already a 3+1 statement: horizon related information is present in the initial data, but it is encoded quasi locally.

13.4 Solving the constraints: conformal geometry as an elliptic engine

A practical initial value problem is not simply "pick any γ_{ij} and K_{ij}." It is "choose a set of free data, then solve the constraints to complete the rest." The York Lichnerowicz conformal decomposition is a standard method. The underlying idea is that the constraints resemble elliptic equations once the right variables are chosen, and elliptic equations are well suited for specifying boundary conditions that encode physical desiderata like asymptotic flatness and desired masses.

One common approach writes the spatial metric as a conformal rescaling

$$\gamma_{ij} = \psi^4 \tilde{\gamma}_{ij}, \tag{13.9}$$

where $\tilde{\gamma}_{ij}$ is a chosen conformal metric and ψ is the conformal factor to be solved for. The extrinsic curvature is decomposed into trace and trace free parts,

$$K_{ij} = A_{ij} + \frac{1}{3}\gamma_{ij}K, \qquad \gamma^{ij}A_{ij} = 0, \tag{13.10}$$

and the trace free part is further split into transverse and longitudinal pieces relative to $\tilde{\gamma}_{ij}$. The upshot is that the Hamiltonian constraint becomes a nonlinear elliptic equation for ψ and the momentum constraint becomes an elliptic system for a vector potential that constructs the longitudinal part of A_{ij}. The details matter for implementation, but the geometric message is simple: *initial data construction is an elliptic geometry problem in disguise* [2, 3].

Phenomenon statement 3 (punctures and excision as geometric boundary choices). In black hole initial data, singular interior regions can be handled by treating them as boundaries (excision) or by allowing the conformal factor to have puncture like singularities that represent additional asymptotic ends. These are not ad hoc tricks: they are choices of boundary conditions and topology for the elliptic constraint system, and they directly affect the global geometry of the initial slice.

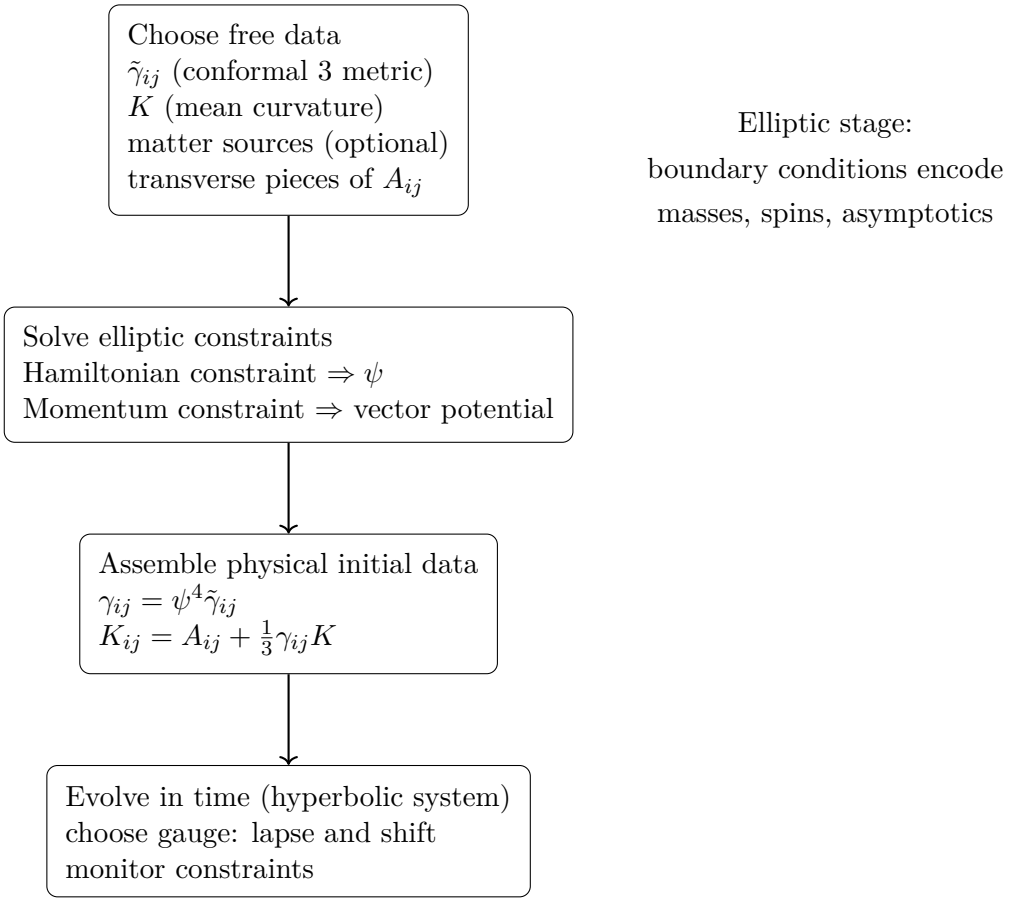


Figure 13.3: **Conceptual diagram (non data).** A typical 3+1 initial data pipeline. Constraint solving is an elliptic geometry step that turns chosen conformal and extrinsic free data into physical (γ_{ij}, K_{ij}), which are then evolved by a chosen hyperbolic evolution system [2–4].

13.4.1 Conceptual diagram: constraint solving pipeline

13.5 Evolution equations and the meaning of gauge choices

Once initial data satisfy the constraints, the remaining components of the Einstein equations provide evolution equations. In the ADM formulation, these evolution equations describe how γ_{ij} and K_{ij} change with time. Schematically,

$$\partial_t \gamma_{ij} = -2\alpha K_{ij} + \mathcal{L}_\beta \gamma_{ij}, \tag{13.11}$$

$$\partial_t K_{ij} = -D_i D_j \alpha + \alpha \left({}^{(3)}R_{ij} + K K_{ij} - 2K_{ik}K^k{}_j \right) + \mathcal{L}_\beta K_{ij} - 8\pi\alpha \left(S_{ij} - \frac{1}{2}\gamma_{ij}(S - \rho) \right), \tag{13.12}$$

where ${}^{(3)}R_{ij}$ is the Ricci tensor of γ_{ij} and $\mathcal{L}_\beta$ denotes Lie derivative along the shift.

Equations (13.11) and (13.12) show something profound: lapse and shift appear as freely specifiable functions that reshape the evolution. They are not determined by the Einstein equations alone. In physical terms, this is diffeomorphism invariance made concrete: you may choose how to label slices and how to drag coordinates, and those choices feed directly into the PDE system.

In practice, straightforward ADM evolution is often numerically unstable for strong field problems. Modern numerical relativity typically uses reformulations such as BSSN, which introduces conformal variables and auxiliary connection functions to improve hyperbolicity and constraint behavior [4, 9]. The geometric core remains the same: evolve a three metric and an extrinsic curvature, with gauge conditions prescribing lapse and shift.

Phenomenon statement 4 (gauge as a shock absorber). In merger simulations, gauge conditions are engineered to prevent coordinate pathologies such as slice stretching and to keep the horizon region numerically well behaved. The fact that you can "tune" the lapse and shift is not cheating; it is using the mathematical freedom of diffeomorphisms to choose a foliation that is well adapted to the physics you want to read off, such as stable horizon tracking or clean waveform extraction.

13.5.1 Conceptual diagram: lapse and shift as a controlled deformation

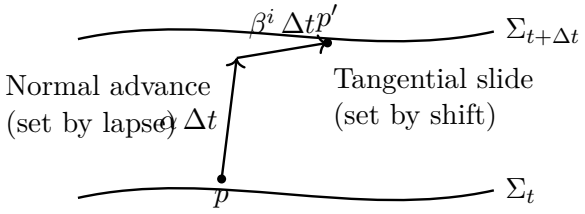


Figure 13.4: **Conceptual diagram (non data).** A coordinate time step Δt maps a point $p \in \Sigma_t$ to a point $p' \in \Sigma_{t+\Delta t}$ via a normal displacement of size $\alpha \Delta t$ followed by a tangential displacement $\beta^i \Delta t$. This is the geometric content of gauge choices in 3+1 evolution [2, 5].

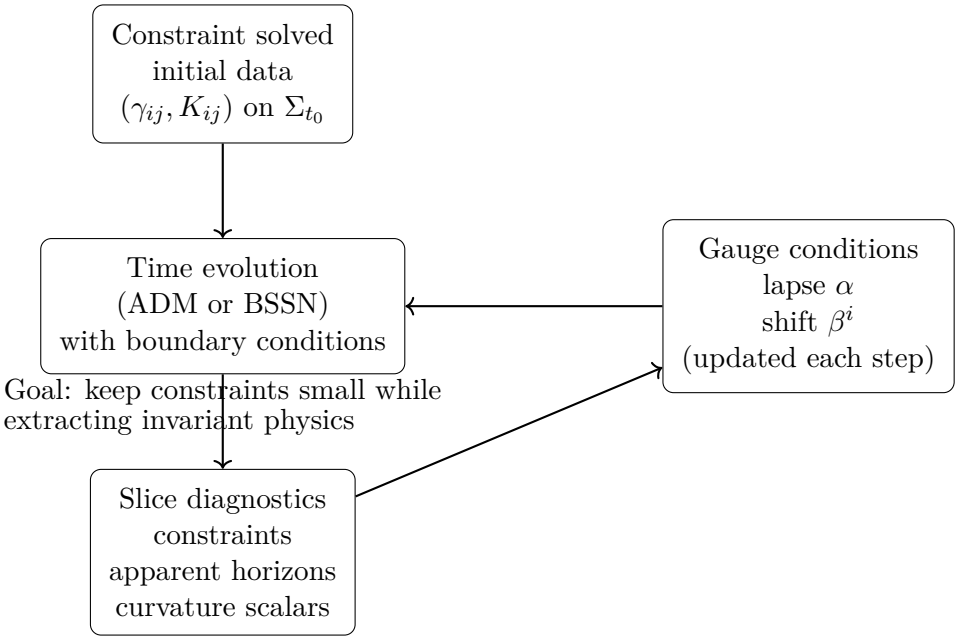


Figure 13.5: **Conceptual diagram (non data).** A typical numerical relativity workflow viewed geometrically. Constraint solving produces initial data, evolution propagates the geometry, diagnostics monitor constraint satisfaction and quasi local structures such as apparent horizons, and gauge conditions update lapse and shift to maintain a well behaved foliation [4, 7, 8].

13.5.2 Conceptual diagram: evolution, gauge, and constraint monitoring loop

13.6 Apparent horizons as quasi local geometry on a slice

A black hole event horizon is a global object: it is defined by causal structure and depends on the entire future of spacetime. In contrast, a 3+1 formulation naturally encounters *apparent horizons*, which are quasi local and can be found on a given slice without knowledge of the entire spacetime. This is critical in time dependent situations such as mergers, where apparent horizon tracking provides a robust way to measure mass and spin of the remnant while the spacetime is still settling.

Let S be a closed two surface embedded in a slice Σ_t, with outward pointing unit spatial normal s^i within the slice. One defines two future directed null normals

$$\ell^a = n^a + s^a, \qquad k^a = n^a - s^a, \tag{13.13}$$

up to normalization conventions. The expansions $\theta_{(\ell)}$ and $\theta_{(k)}$ measure how bundles of

null geodesics orthogonal to S diverge or converge. An apparent horizon on the slice is typically defined by

$$\theta_{(\ell)} = 0, \qquad \theta_{(k)} < 0. \tag{13.14}$$

The condition $\theta_{(\ell)} = 0$ becomes a nonlinear elliptic equation on the surface S when expressed in terms of γ_{ij} and K_{ij}. That is why apparent horizon finding is often described as solving a geometric PDE problem [6].

13.6.1 Conceptual diagram: apparent horizon finding on a slice

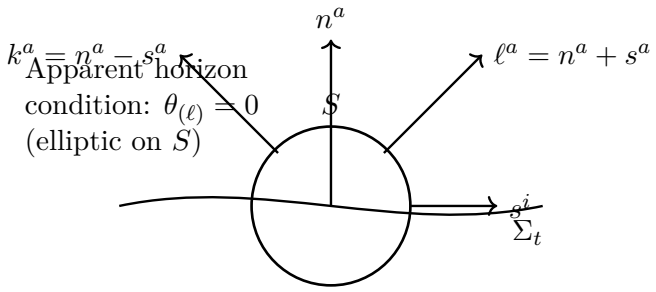


Figure 13.6: **Conceptual diagram (non data).** A closed surface S in a slice Σ_t becomes an apparent horizon when the outward null expansion vanishes. In 3+1 terms, this is a nonlinear elliptic condition involving γ_{ij} and K_{ij}, making horizon finding a geometric PDE problem [6].

13.7 Data based illustrations: catalog level shadows of 3+1 geometry

The previous sections developed the 3+1 split as a geometric re expression of Einstein dynamics. To make the connection tangible, we now use a real world dataset: parameter summaries for a subset of events in the GWTC 3 catalog, provided openly by the Gravitational Wave Open Science Center (GWOSC) event portal [10, 11]. The GWTC 3 confident list contains 35 events from the O3b observing run. From the event portal release table we extract median values for a representative subset of 22 events from the GWTC 3 confident release and use them as a compact dataset for plots. We then compute two derived quantities that are directly meaningful for horizon geometry:

- the mass ratio $q = m_1/m_2$ (a dimensionless measure of initial data asymmetry),
- the radiated mass fraction $f_{\rm rad} = 1 - M_{\rm final}/M_{\rm tot}$ (a proxy for how much energy left the system as gravitational radiation, and therefore how much the horizon mass decreased from total to remnant).

These derived quantities are simple algebraic functions of catalog parameters, so the figures remain strictly dataset based while still illustrating geometric consequences.

13.7.1 Dataset provenance and parameter meanings

The catalog table provides "Mass 1" and "Mass 2" in the source frame, the network matched filter signal to noise ratio (SNR), luminosity distance in megaparsecs, the effective inspiral spin parameter χ_{eff}, the total source frame mass, a chirp mass, a redshift, the false alarm rate (FAR), the astrophysical probability p_{astro}, and the final mass. GWOSC provides definitions and schema information for these parameters, including χ_{eff} and source frame remnant properties [12, 13].

Phenomenon statement 5 (why catalog parameters are "initial data projections"). Catalog numbers like m_1, m_2, and χ_{eff} are not the initial data themselves. The initial data in 3+1 form are (γ_{ij}, K_{ij}) on a slice. However, for an isolated binary far from merger, those initial data have gauge invariant summaries that behave like masses and spins. The catalog parameters are therefore best understood as low dimensional projections of a high dimensional geometric initial state.

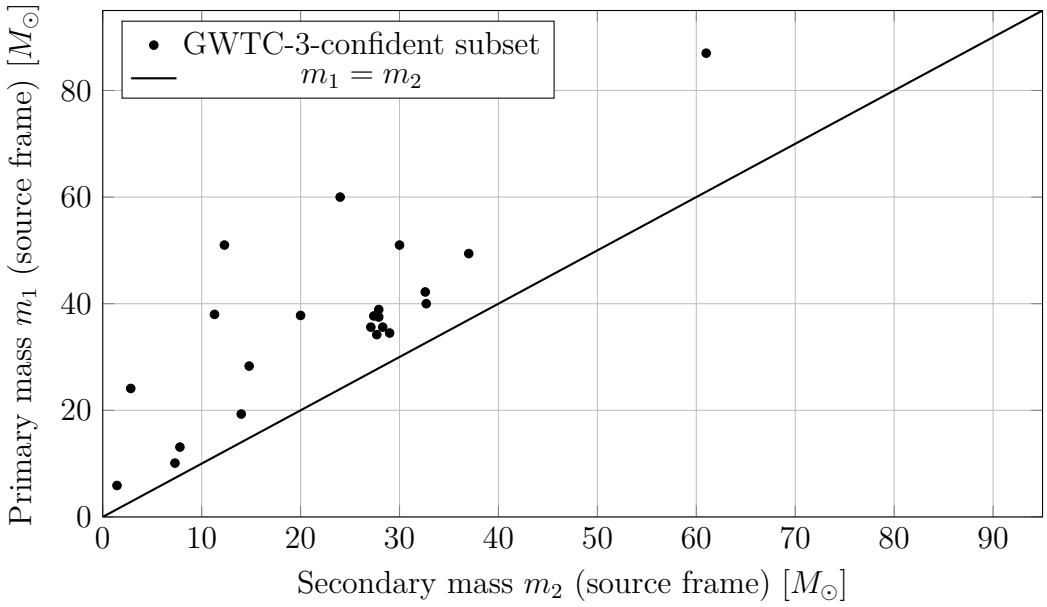


Figure 13.7: **Data based Figure 1.** Component mass plane for 22 GWTC 3 confident events (medians). Points above the diagonal correspond to asymmetric binaries with $m_1 > m_2$. In a 3+1 initial data construction, the masses influence the conformal factor and extrinsic curvature content needed to represent the binary in a slice geometry. Data source: GWOSC event portal release table [10, 11].

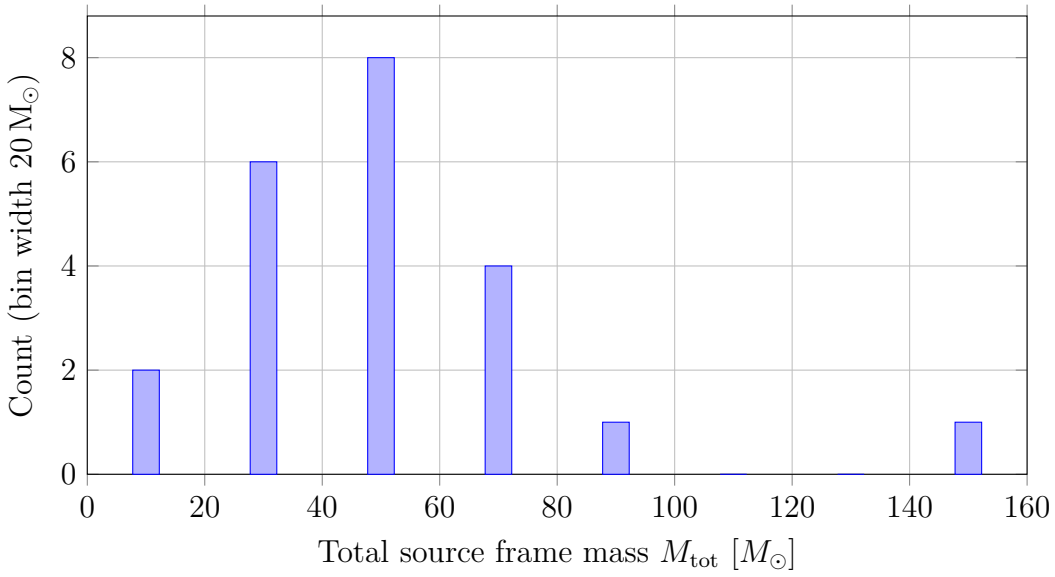


Figure 13.8: **Data based Figure 2.** Histogram of total source frame mass for the same dataset. The distribution is shaped by astrophysical populations and by selection effects: heavier systems are visible to larger distances but spend fewer cycles in band. In 3+1 language, higher total mass pushes stronger curvature and shorter dynamical time scales into the slice geometry. Data source: GWOSC event portal release table [11].

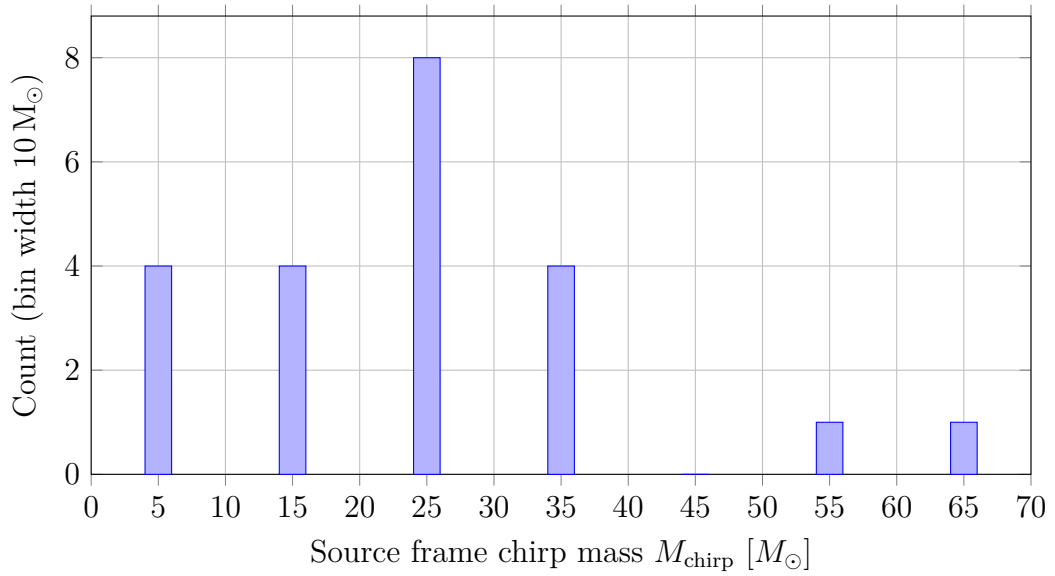


Figure 13.9: **Data based Figure 3.** Histogram of chirp mass M_{chirp}, a combination of m_1 and m_2 that controls leading order inspiral phasing. Chirp mass is a clean bridge between data and geometry: it sets the time scale of orbital frequency growth, which in a 3+1 evolution corresponds to how rapidly the slice geometry changes as the binary tightens. Data source: GWOSC event portal release table [11, 13].

13.7.2 Interpreting the dataset through a 3+1 lens

Even without solving a single constraint equation explicitly, the dataset suggests several geometric lessons that align with the theory developed earlier.

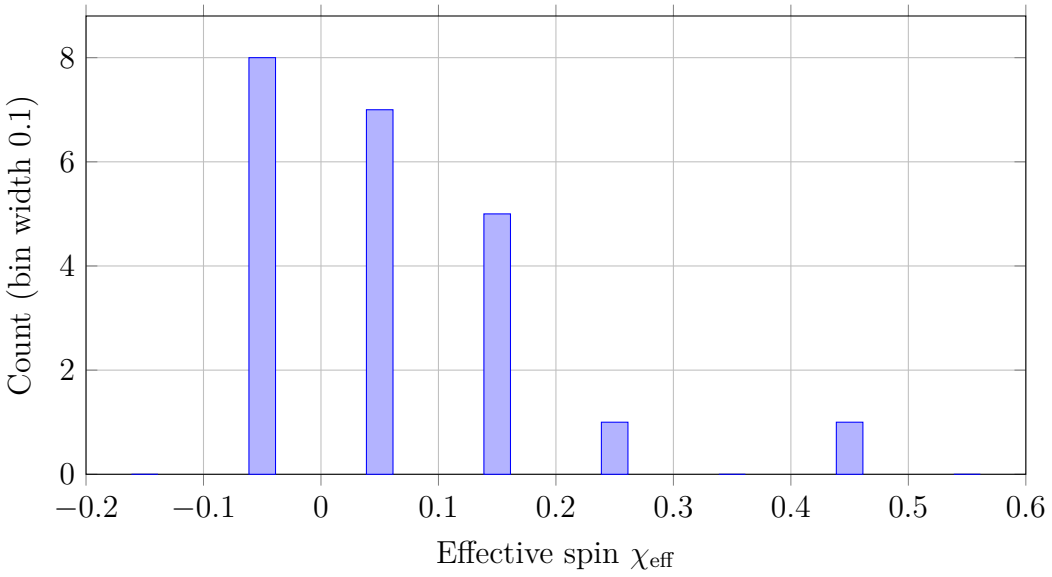


Figure 13.10: **Data based Figure 4.** Histogram of effective aligned spin χ_{eff} (medians). In 3+1 initial data, spin information is encoded primarily through the extrinsic curvature and through the horizon geometry of individual punctures or excision surfaces. The distribution here shows a mixture of positive and negative values, highlighting that the slice embedding data must accommodate both spin aligned and anti aligned configurations. Data source: GWOSC event portal release table and parameter definitions [11, 13].

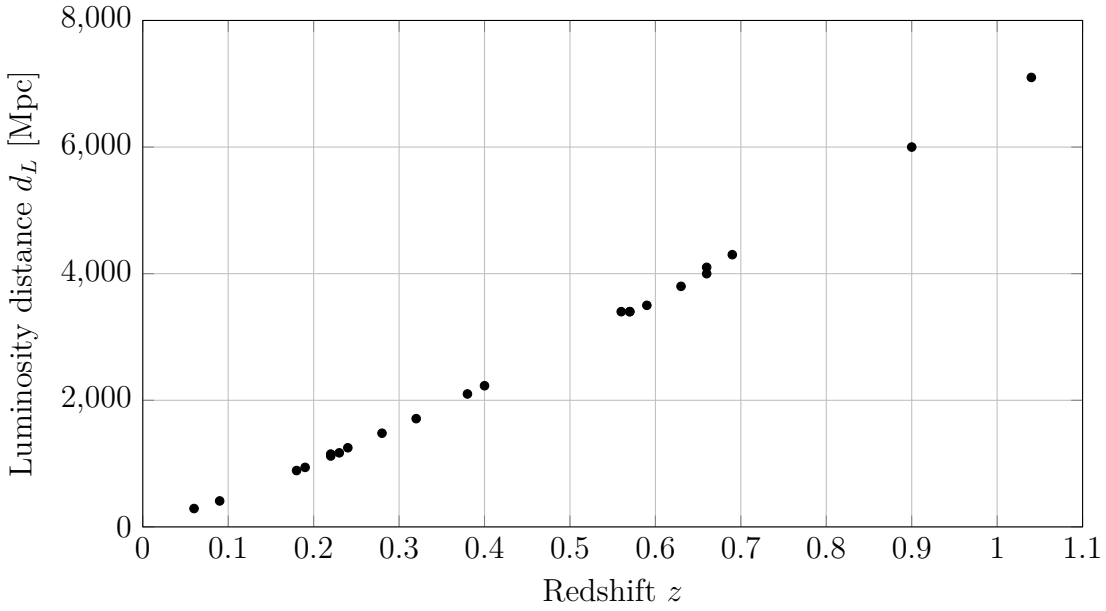


Figure 13.11: **Data based Figure 5.** Luminosity distance versus redshift for the dataset. Although cosmology is not the topic of this chapter, this plot is a reminder that 3+1 simulations are used to produce waveforms, and waveforms are used to infer distances and redshifts. The geometric initial value problem therefore connects, indirectly but concretely, to cosmological scale measurements. Data source: GWOSC event portal release table [11].

In this subset, the primary mass ranges from $m_1 \approx 5.9\,\mathrm{M}_\odot$ to $m_1 \approx 87\,\mathrm{M}_\odot$, the secondary mass ranges from $m_2 \approx 1.44\,\mathrm{M}_\odot$ to $m_2 \approx 61\,\mathrm{M}_\odot$, and the luminosity distance spans

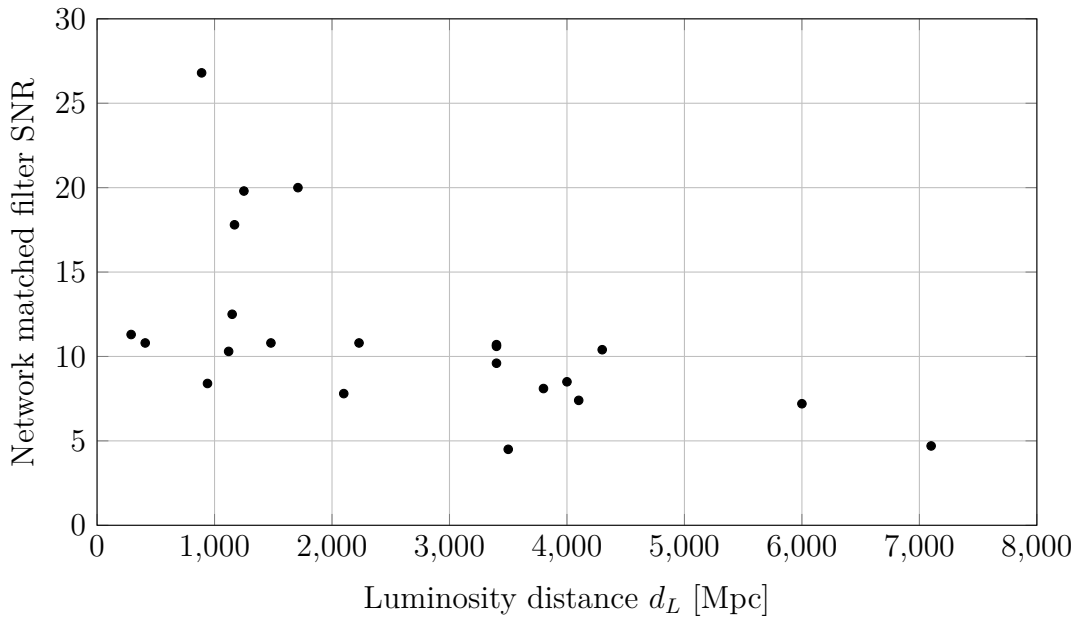


Figure 13.12: **Data based Figure 6.** Detection strength (network SNR) versus distance. The scatter reflects orientation, detector sensitivity at the time of the event, and intrinsic source properties such as mass and spin. In the 3+1 picture, intrinsic properties correspond to different initial data, but the mapping from those data to the observed SNR passes through evolution, wave extraction, and the detector response. Data source: GWOSC event portal release table [11].

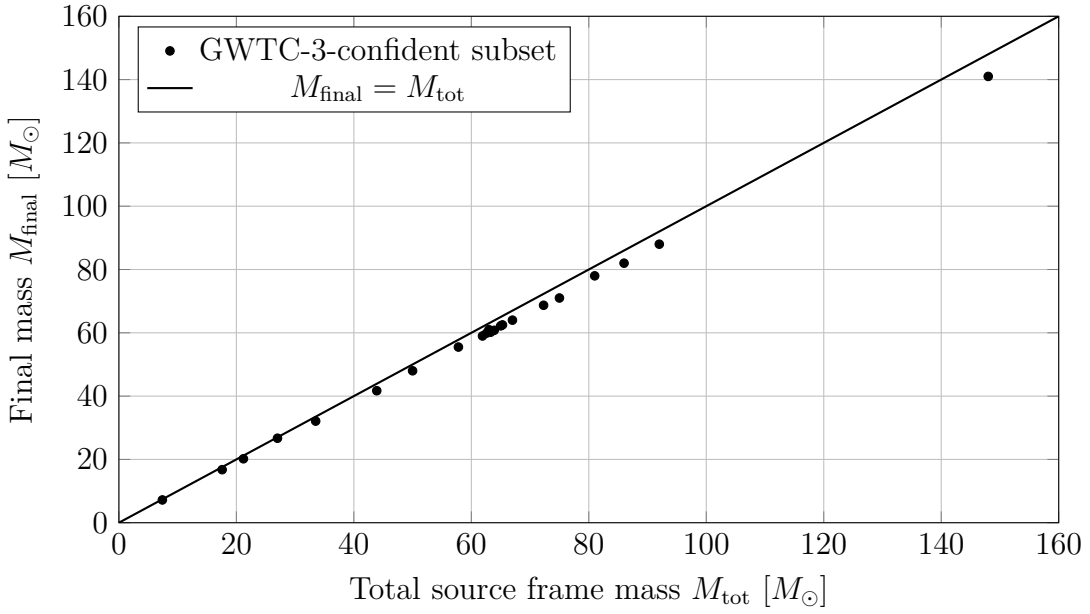


Figure 13.13: **Data based Figure 7.** Final mass versus total mass. Points lie below the equality line because a fraction of the initial mass energy is carried away by gravitational radiation. In 3+1 simulations, this radiated energy is reflected geometrically by changes in horizon area and by the flux of curvature information through large extraction surfaces. Data source: GWOSC event portal release table [11].

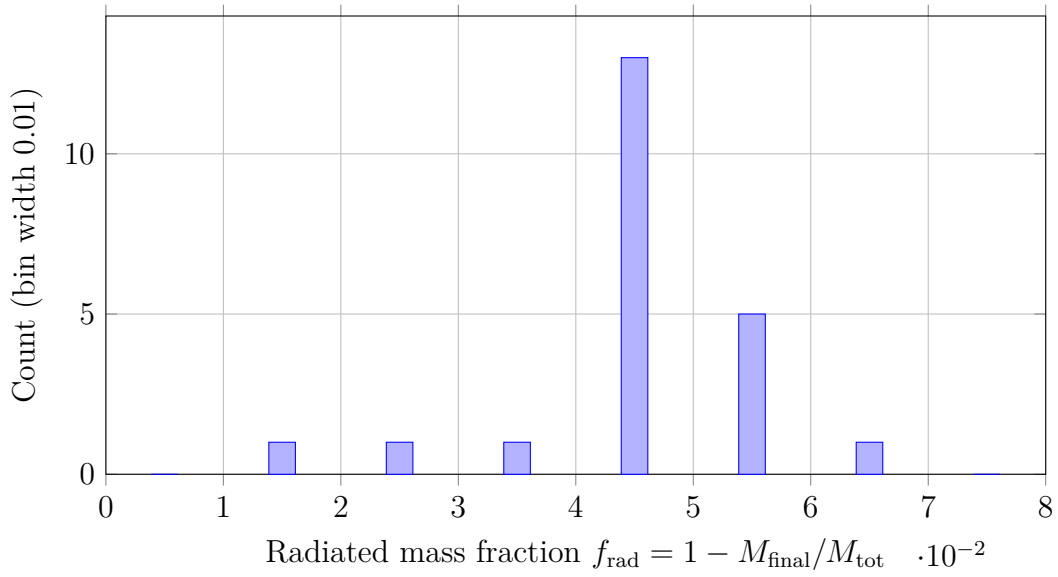


Figure 13.14: **Data based Figure 8.** Distribution of the radiated mass fraction computed from catalog medians. This is an example of mathematical modification of an open dataset: we form f_{rad} directly from $(M_{\mathrm{tot}}, M_{\mathrm{final}})$. In this subset, f_{rad} is typically a few percent, consistent with the expectation that mergers radiate a noticeable but not dominant fraction of the total mass energy. In geometric terms, this fraction is the difference between initial mass content on the slice and the remnant horizon mass. Data source: GWOSC event portal release table [11].

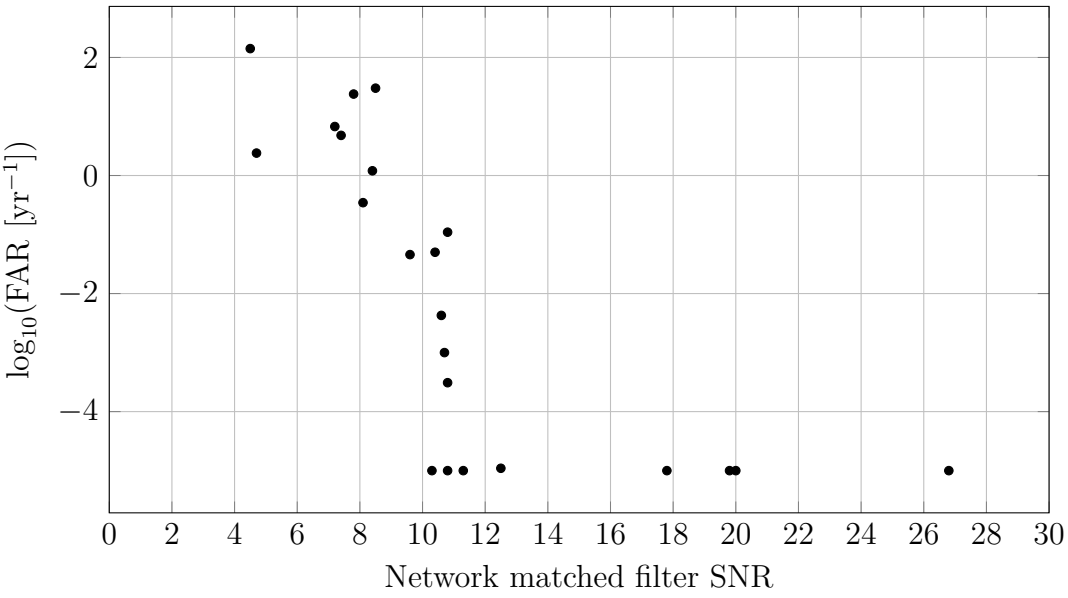


Figure 13.15: **Data based Figure 9.** Log false alarm rate (FAR) versus SNR. FAR is a data analysis quantity, but it illustrates a practical point about the initial value problem: a model that violates constraints or mishandles gauge can produce inaccurate waveforms, which in turn can degrade detection statistics. Conversely, robust 3+1 modeling supports accurate templates and reliable significance measures. For events listed with FAR upper bounds, the bound is used as a numeric value for plotting, so the corresponding points should be interpreted as approximate. Data source: GWOSC event portal release table [11].

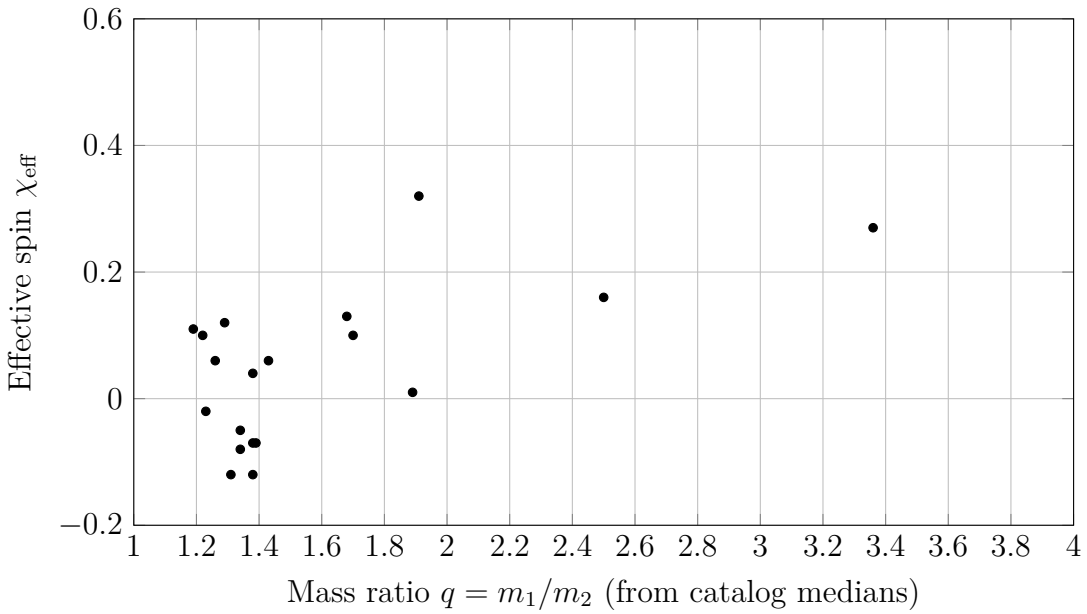


Figure 13.16: **Data based Figure 10.** Effective spin versus mass ratio for the dataset (derived q). In initial data language, mass ratio controls the asymmetry of the geometry and often the difficulty of maintaining clean gauge behavior near both objects. Spin affects extrinsic curvature content and can distort horizons, altering the quasi local geometry that apparent horizon finders track during evolution. Data source: GWOSC event portal release table [11, 13].

roughly 290 Mpc to 7100 Mpc. The redshifts range from $z \approx 0.06$ to $z \approx 1.04$. The radiated mass fraction computed from catalog medians spans approximately $f_{\rm rad} \approx 0.011$ to $f_{\rm rad} \approx 0.054$, with a mean near ≈ 0.042. These simple descriptive numbers already hint at the diversity of geometric initial data that a 3+1 formulation must accommodate.

First, the mass plane (Data Figure 1) shows a wide range of component masses. In a conformal decomposition, changing m_1 and m_2 is not merely rescaling; it changes the way curvature is distributed across the slice and the way the extrinsic curvature must encode orbital motion and spin. High mass systems have shorter geometric time scales near merger, meaning a given slicing condition has less "time" to smooth out coordinate distortions before the merger happens. This is one reason gauge design is so central in practice.

Second, the chirp mass histogram (Data Figure 3) clusters many events around intermediate values. Chirp mass is the combination of masses that most cleanly controls inspiral phasing, which in 3+1 terms is the rate of change of orbital curvature. The fact that many events share similar chirp masses suggests why waveform models can be built from families of simulations: the geometric initial data vary in a structured way across parameter space.

Third, the effective spin distribution (Data Figure 4) is centered near small values with tails extending positive and negative. This is a reminder that "spin" in a catalog is not a

separate add on; it is part of the geometric initial data, encoded through K_{ij} and through the horizon geometry of each object. In 3+1 evolution, spin also affects frame dragging, which changes both the physical dynamics and the coordinate behavior for a given gauge choice.

Fourth, the remnant mass relation (Data Figure 7) and the radiated fraction (Data Figure 8) emphasize that horizon geometry is not static. The remnant horizon mass is smaller than the total source mass because energy has been carried away. In 3+1 computations, one often measures remnant properties by locating the apparent horizon and computing quasi local invariants. The dataset is therefore a catalog level reflection of apparent horizon tracking: a horizon is found, its area and spin are estimated, and those estimates become part of the event summary.

Phenomenon statement 6 (why apparent horizons are operationally important). During a merger, the event horizon is not practically available because it requires the entire future. Apparent horizons are available slice by slice. They provide the remnant mass and spin that appear in observational catalogs. Therefore, apparent horizon finding is not a niche numerical technique; it is part of the conceptual bridge between 3+1 geometry and real observational inferences.

13.8 Practical geometric pitfalls: what can go wrong, and why it is geometric

Because the 3+1 decomposition introduces gauge freedom, it also introduces the possibility of choosing a foliation that behaves poorly. A classic pathology is slice stretching near horizons: if the lapse collapses too slowly or the shift fails to move coordinates appropriately, grid points can pile up and the numerical representation of geometry becomes distorted. In geometric language, the foliation is threading through spacetime in a way that produces large gradients in γ_{ij} and K_{ij} that are not physically required but are coordinate artifacts.

Another pitfall is constraint drift. Even if the constraints are satisfied initially, numerical errors can introduce small violations that, if not controlled, grow and contaminate the solution. Many modern formulations are designed so that constraint violations propagate away or are damped. This is again geometric: it is about the PDE character of the chosen representation of curvature and embedding.

Boundary conditions are a third subtlety. In an asymptotically flat problem one wants the slice geometry to approach the Minkowski metric at large distances, and one wants wave like perturbations to leave the domain without reflection. But in a finite computational

domain, boundaries are artificial hypersurfaces that must be given conditions. The quality of those conditions affects waveform extraction, and therefore affects data analysis quantities like SNR and parameter inference. The connection from boundary geometry to observational catalog entries is not metaphorical; it is causal.

13.9 Conclusion

The 3+1 decomposition is the place where general relativity becomes an initial value theory without ceasing to be geometry. A foliation introduces a controlled way to talk about "now" while respecting coordinate freedom. The induced metric γ_{ij} and extrinsic curvature K_{ij} are not arbitrary fields: they are intrinsic and extrinsic geometry, linked by constraints that encode whether a slice can be embedded in a spacetime satisfying Einstein's equations. Gauss Codazzi relations explain why those constraints exist, and conformal methods explain how they can be solved in practice. Evolution equations then deform the slice through spacetime, with lapse and shift acting as geometric gauge controls rather than physical forces.

Horizons, especially apparent horizons, sit naturally inside this framework. They are quasi local geometric structures that can be found on a slice and tracked through time. This makes them essential for interpreting dynamical black hole spacetimes and for connecting simulations and theory to observables.

Finally, the chapter grounded these concepts with a real dataset from the GWTC 3 confident catalog. By plotting masses, effective spin, distance, significance measures, and remnant mass, and by computing derived quantities such as mass ratio and radiated fraction, we showed how high level geometric ideas have measurable footprints. The data do not replace the theory, but they do act as a reality check: the initial value problem is not only a mathematical formulation, it is the backbone behind the observational era of black hole dynamics.

Bibliography

[1] R. Arnowitt, S. Deser, and C. W. Misner, "The Dynamics of General Relativity," arXiv:gr-qc/0405109, 2004. Available: https://arxiv.org/abs/gr-qc/0405109

[2] E. Gourgoulhon, "3+1 Formalism and Bases of Numerical Relativity," arXiv:gr-qc/0703035, 2007. Available: https://arxiv.org/abs/gr-qc/0703035

[3] J. W. York Jr., "Role of Conformal Three Geometry in the Dynamics of Gravitation," *Physical Review Letters*, vol. 28, pp. 1082-1085, 1972. Available: https://doi.org/10.1103/PhysRevLett.28.1082

[4] T. W. Baumgarte and S. L. Shapiro, "On the Numerical Integration of Einstein's Field Equations," *Physical Review D*, vol. 59, 024007, 1998.

[5] M. Alcubierre, *Introduction to 3+1 Numerical Relativity*. Oxford, UK: Oxford University Press, 2008. Available: https://global.oup.com/academic/product/introduction-to-31-numerical-relativity-9780199205677

[6] J. Thornburg, "Finding Apparent Horizons in Numerical Relativity," *Living Reviews in Relativity*, vol. 10, no. 3, 2007.

[7] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, "Accurate Evolutions of Orbiting Black Hole Binaries without Excision," *Physical Review Letters*, vol. 96, 111101, 2006. Available: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.96.111101

[8] J. G. Baker *et al.*, "Gravitational Wave Extraction from an Inspiraling Configuration of Merging Black Holes," *Physical Review Letters*, vol. 96, 111102, 2006.

[9] M. Shibata and T. Nakamura, "Evolution of Three Dimensional Gravitational Waves: Harmonic Slicing Case," *Physical Review D*, vol. 52, pp. 5428-5444, 1995.

[10] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run," arXiv:2111.03606, 2021. Available: https://arxiv.org/abs/2111.03606

[11] Gravitational Wave Open Science Center (GWOSC), "GWTC-3-confident: Confident events from the O3b observing run for LIGO and Virgo," Event Portal release table, . Available: https://gwosc.org/eventapi/html/GWTC-3-confident/

[12] Gravitational Wave Open Science Center (GWOSC), "Event Portal Usage Notes," . Available: https://gwosc.org/eventportaldocs/

[13] GWOSC Tutorial Project, "GWOSC Catalog Schema and Parameter Definitions," . Available: https://gwosc-tutorial.github.io/gwosc-catalog/schema/

[14] Gravitational Wave Open Science Center (GWOSC), "GWOSC Archive," . Available: https://gwosc.org/archive/

Black Holes Geometry: Curvature and Horizons

Chapter 14: Tetrads, Spinors, and Newman-Penrose Geometry

Contents

IV Geometric Methods for Dynamics and Perturbations **1**

14 Tetrads, Spinors, and Newman-Penrose Geometry **2**

Abstract . . . 2
Keywords . . . 2
14.1 Chapter roadmap and index table of contents . . . 2
14.2 Introduction: why frames and spinors matter for black holes . . . 3
14.3 Orthonormal tetrads and measured geometry . . . 4
14.3.1 Frames, coframes, and local Minkowski structure . . . 4
14.3.2 Connection 1-forms and Cartan structure equations . . . 5
14.4 Null tetrads and spin coefficients . . . 5
14.4.1 From orthonormal tetrads to null tetrads . . . 5
14.4.2 Directional derivatives and spin coefficients . . . 6
14.5 Two-spinors as square roots of null geometry . . . 7
14.5.1 Spin spaces and index structure . . . 7
14.5.2 Spin frames and the null tetrad correspondence . . . 7
14.6 Weyl scalars as curvature multipole moments . . . 8
14.6.1 Vacuum curvature and the Weyl tensor . . . 8
14.6.2 From Ψ_4 to what detectors measure . . . 8
14.7 Petrov classification and Kerr specialness . . . 9
14.7.1 Algebraic types as statements about principal null directions . . . 9
14.7.2 Why Kerr specialness becomes calculational power . . . 9
14.8 Data-based illustrations: catalog-level curvature and radiation scales . . . 10
14.8.1 Mass ratios and binary asymmetry in the data . . . 11
14.8.2 Spin information and why $\chi_{\rm eff}$ is a geometric compression . . . 11
14.8.3 Distance, signal-to-noise, and detector tetrads . . . 11
14.8.4 Redshift and total mass . . . 14
14.8.5 Distribution of remnant masses . . . 14
14.8.6 Data-derived ringdown scale from remnant mass . . . 14
14.9 Applications and phenomenon statements grounded in real data . . . 14

14.9.1 Phenomenon 1: detector polarizations as tetrad projections . . . 14
14.9.2 Phenomenon 2: ringdown spectroscopy as curvature eigenmodes 15
14.9.3 Phenomenon 3: stellar-mass black holes in binaries as horizon laboratories . . . 16
14.9.4 Phenomenon 4: horizon-scale imaging and the universality of the ring diameter . . . 18
14.9.5 Phenomenon 5: why NP curvature scalars are useful in numerical and data pipelines . . . 18
14.10Conclusion . . . 19
Manual index of key terms . . . 19

List of Figures

14.1 Conceptual diagram: coordinate grid versus an orthonormal tetrad. The tetrad vectors represent locally measured directions. 5
14.2 Conceptual diagram: the null tetrad $(\ell, n, m, \bar{m})$ and its geometric meaning. 6
14.3 Conceptual diagram: NP operators and the interpretation of spin coefficients as geometric optics data. 7
14.4 Conceptual diagram: spinor dyad, soldering form, and induced null tetrad. 8
14.5 Conceptual diagram: peeling picture of Weyl scalars and asymptotic radiation. 9
14.6 Conceptual diagram: Petrov classification intuition using NP Weyl scalars. 10
14.7 Data-based figure: source-frame component masses for GWTC-3 confident events. Values are medians from the GWOSC table [6]. 11
14.8 Data-based figure: effective spin versus chirp mass for GWTC-3 confident events (medians from GWOSC) [6]. 12
14.9 Data-based figure: network SNR versus luminosity distance for GWTC-3 confident events [6]. The curve is a guide-to-eye inverse scaling, not a fit. 12
14.10 Data-based figure: total mass versus redshift for GWTC-3 confident events [6]. 13
14.11 Data-based figure: histogram of final (remnant) masses in the GWTC-3 confident sample (medians from GWOSC) [6]. 13
14.12 Data-based figure: ringdown frequency scale computed from GWTC-3 remnant masses using Eq. (14.21). Data: GWOSC [6]. Constant source: standard QNM results summarized in [5]. 14
14.13 Data-based figure: black hole mass versus orbital period for a compiled X-ray binary sample. Data source: stellarcollapse.org observed black hole masses [8]. 16
14.14 Data-based figure: companion mass versus black hole mass for the same X-ray binary compilation [8]. 17
14.15 Data-based figure: histogram of stellar-mass black hole masses in the compiled X-ray binary sample [8]. 17

14.16Data-based figure: EHT observed ring diameters compared to a Schwarzschild shadow diameter prediction computed from masses and distances quoted in EHT publications [9, 10]. 18

List of Tables

14.1 Index table of contents for Chapter 14: where key ideas and the data-based figures appear. 3

Part IV

Geometric Methods for Dynamics and Perturbations

Chapter 14

Tetrads, Spinors, and Newman-Penrose Geometry

Abstract: Tetrads and spinors repackage Lorentzian geometry into objects that behave like measured quantities: what an ideal observer can resolve, and what a null ray carries. In black hole spacetimes, that repackaging is the gateway to the Newman-Penrose calculus, where connection coefficients become physically interpretable optical scalars and curvature decomposes into Weyl scalars that separate Coulomb-like curvature from radiative degrees of freedom. This chapter develops orthonormal frames and null tetrads, introduces two-spinor language as the square root of null directions, and builds the Newman-Penrose dictionary that turns covariant derivatives into spin coefficients and the Weyl tensor into five complex scalars. To connect the formalism to the real world, we analyze open gravitational-wave catalog data and black hole mass measurements and derive observationally anchored curvature and ringdown scales, and we connect horizon geometry to event-horizon-scale imaging.

Keywords: tetrads, orthonormal frames, null tetrads, spin coefficients, two-spinors, Weyl scalars, Petrov classification, Kerr type D, Teukolsky equation, gravitational waves, ringdown spectroscopy, event horizon imaging.

14.1 Chapter roadmap and index table of contents

This chapter introduces a new language while insisting that the language remains physically transparent. The recommended reading strategy is: learn what each object means, then learn how it transforms, and only then lean into the algebra. Table 14.1 acts as an index-style roadmap linking the main conceptual milestones to the places where they are defined, used, and connected to observation.

Table 14.1: Index table of contents for Chapter 14: where key ideas and the data-based figures appear.

Theme	Where it is built	Where it is applied (figures)
Orthonormal tetrads, coframes, connection 1-forms	Sections 14.3-14.3.2	Conceptual Figs. 14.1, 14.2
Null tetrads and optical scalars	Sections 14.4-14.4.2	Conceptual Figs. 14.3, 14.5
Two-spinors and spin frames	Sections 14.5-14.5.2	Conceptual Fig. 14.4
Weyl scalars and radiation content	Sections 14.6-14.6.2	Data Figs. 14.7-14.11
Petrov classification and Kerr specialness	Sections 14.7-14.7.2	Conceptual Fig. 14.6; Data Fig. 14.12
Real-world horizons and observations	Section 14.9	Data Figs. 14.13-14.16

14.2 Introduction: why frames and spinors matter for black holes

Black hole geometry invites a recurring mistake: treating coordinates as reality. Coordinates are useful, but they are not what an observer measures, and they are not what a null ray transports. A black hole horizon is not only a surface described by an equation like $r = 2M$ or $\Delta = 0$. It is also a causal boundary whose physics is carried by null directions and whose geometry is encoded in how families of null geodesics expand, shear, and twist. The most natural language for that is not coordinate components, but frames adapted to physical directions.

A tetrad is a choice of four linearly independent vectors at each spacetime point. When the tetrad is orthonormal, it is a local surrogate for an ideal laboratory: it tells you what a small accelerometer, gyroscope, or photon wave-vector looks like in a locally Minkowskian measurement apparatus. This idea becomes sharper in Lorentzian geometry because light selects a preferred structure: the light cone. When we reorganize an orthonormal tetrad into a null tetrad, we gain a calculus whose basis vectors point along the very directions that define horizons and gravitational radiation.

Spinors enter because null directions are not merely a subset of vectors. In four dimensions, they have a square-root structure. Two-spinors generate null vectors and null bivectors, making them natural objects for describing principal null directions and for classifying curvature algebraically. This matters in black hole spacetimes because Kerr is not merely a vacuum solution. It is algebraically special (Petrov type D), and

that specialness is what makes its perturbations separable and its radiation describable by decoupled scalar equations [1, 3, 4].

The Newman-Penrose (NP) formalism ties these threads together. It replaces Christoffel symbols with spin coefficients that behave like geometric optics data, and it replaces the full Weyl tensor with five complex scalars $\Psi_0, \ldots, \Psi_4$ that behave like curvature multipole moments [1]. Once that dictionary is internalized, several later chapters stop wobbling. Perturbation theory, optical geometry, and wave extraction in numerical and observational contexts become more transparent because the formalism keeps null structure explicit.

To keep the chapter grounded, we repeatedly connect the formalism to open datasets: gravitational-wave catalog parameters from the GWOSC GWTC-3 confident table [6], stellar-mass black hole masses in X-ray binaries compiled at stellarcollapse.org [8], and event-horizon-scale ring diameters measured by the Event Horizon Telescope (EHT) for M87 and Sagittarius A [9, 10]. These datasets do not list tetrads explicitly, but they determine curvature scales and horizon scales. That is the point: tetrads and NP scalars are the geometric infrastructure that makes the data physically interpretable.

14.3 Orthonormal tetrads and measured geometry

14.3.1 Frames, coframes, and local Minkowski structure

A tetrad (or vierbein) is a set of four vectors $\{e_{(a)}\}$ defined at each spacetime point, forming a basis of the tangent space. The parenthesized index (a) is a frame index. It is not a coordinate index and it does not transform under coordinate changes in the usual tensorial way. Instead, it transforms under local Lorentz transformations. This is the first conceptual pivot: separate the freedom of choosing coordinates from the freedom of choosing a local physical frame.

Let g_{ab} be the spacetime metric with signature $(-,+,+,+)$. An orthonormal tetrad satisfies

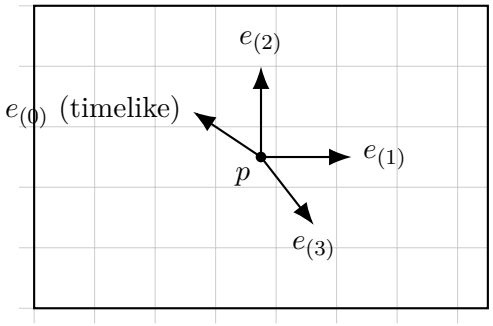
$$g\big(e_{(a)}, e_{(b)}\big) = \eta_{(a)(b)}, \qquad \eta_{(a)(b)} = \mathrm{diag}(-1,1,1,1). \tag{14.1}$$

The dual basis (coframe) $\{\theta^{(a)}\}$ satisfies $\theta^{(a)}(e_{(b)}) = \delta^{(a)}_{(b)}$, and the metric can be written as

$$g = \eta_{(a)(b)}\, \theta^{(a)} \otimes \theta^{(b)}. \tag{14.2}$$

This formula says: the metric is Minkowski in the frame basis, and curved geometry is hidden in how the coframe varies from point to point.

Conceptual diagram: coordinates are a chart, tetrads are a local lab



A coordinate basis follows the grid; an orthonormal tetrad follows local physics.

Figure 14.1: Conceptual diagram: coordinate grid versus an orthonormal tetrad. The tetrad vectors represent locally measured directions.

14.3.2 Connection 1-forms and Cartan structure equations

Once we choose a coframe, the Levi-Civita connection is described by connection 1-forms $\omega^{(a)}{}_{(b)}$ defined by

$$\nabla e_{(b)} = \omega^{(a)}{}_{(b)}\, e_{(a)}, \qquad \omega_{(a)(b)} = -\omega_{(b)(a)}. \tag{14.3}$$

The torsion-free condition and the curvature follow from Cartan's structure equations:

$$\mathrm{d}\theta^{(a)} + \omega^{(a)}{}_{(b)} \wedge \theta^{(b)} = 0, \tag{14.4}$$

$$\Omega^{(a)}{}_{(b)} \equiv \mathrm{d}\omega^{(a)}{}_{(b)} + \omega^{(a)}{}_{(c)} \wedge \omega^{(c)}{}_{(b)} = \frac{1}{2} R^{(a)}{}_{(b)(c)(d)}\, \theta^{(c)} \wedge \theta^{(d)}. \tag{14.5}$$

In black hole spacetimes, a tetrad adapted to physically meaningful observers makes redshift, frame dragging, and tidal curvature more explicit than raw coordinate components. When we later reorganize into a null tetrad, these same structure equations become the NP equations that govern spin coefficients and Weyl scalars [1].

14.4 Null tetrads and spin coefficients

14.4.1 From orthonormal tetrads to null tetrads

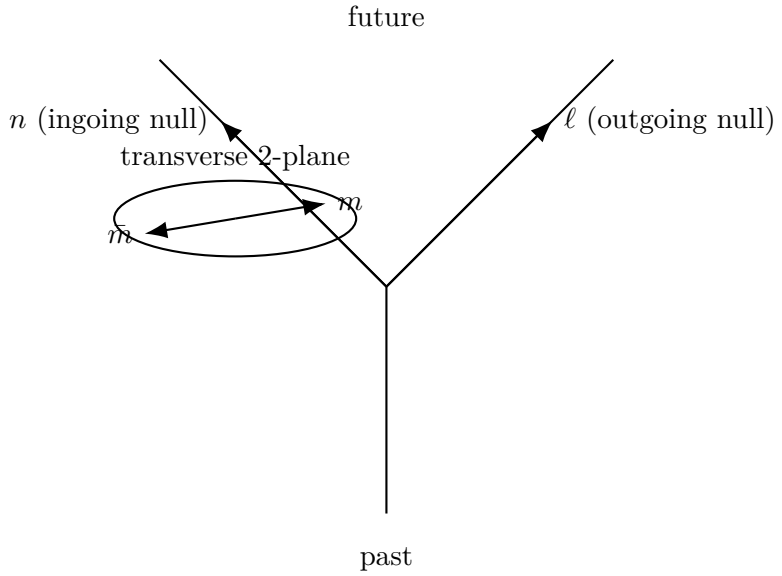
A null tetrad is a set of four vectors $(\ell^a, n^a, m^a, \bar{m}^a)$ where ℓ^a and n^a are real null vectors and m^a is a complex null vector tangent to a spacelike 2-plane. A common normalization is

$$\ell \cdot n = -1, \qquad m \cdot \bar{m} = 1, \qquad \text{all other inner products vanish.} \tag{14.6}$$

Given an orthonormal tetrad $(e_{(0)}, e_{(1)}, e_{(2)}, e_{(3)})$, a standard construction is

$$\ell = \frac{1}{\sqrt{2}}\left(e_{(0)} + e_{(1)}\right), \qquad n = \frac{1}{\sqrt{2}}\left(e_{(0)} - e_{(1)}\right), \qquad m = \frac{1}{\sqrt{2}}\left(e_{(2)} + \mathrm{i}e_{(3)}\right). \quad (14.7)$$

Horizons are null hypersurfaces generated by null curves, so a null tetrad makes the horizon generators and their transverse deformations geometrically explicit.



Conceptual diagram: null tetrad adapted to the light cone and a transverse angular plane.

Figure 14.2: Conceptual diagram: the null tetrad $(\ell, n, m, \bar{m})$ and its geometric meaning.

14.4.2 Directional derivatives and spin coefficients

In the NP formalism, replace coordinate derivatives by directional derivatives along tetrad legs:

$$D \equiv \ell^a \nabla_a, \qquad \Delta \equiv n^a \nabla_a, \qquad \delta \equiv m^a \nabla_a, \qquad \bar{\delta} \equiv \bar{m}^a \nabla_a. \quad (14.8)$$

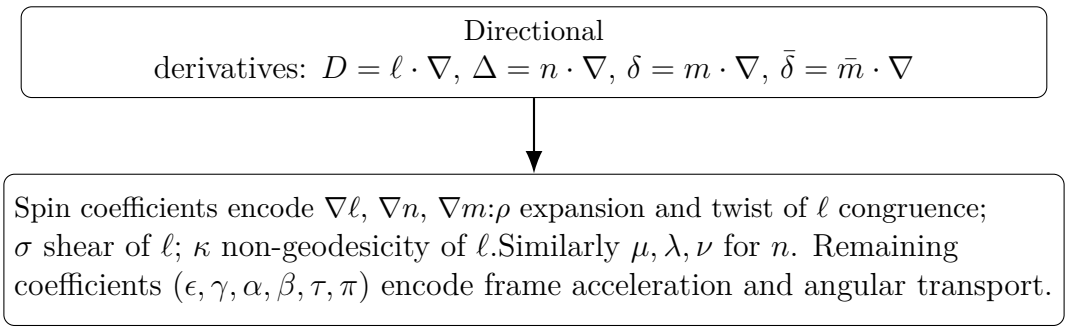
The covariant derivatives of the tetrad vectors are encoded in 12 complex spin coefficients (equivalently 24 real functions). A representative subset is

$$\rho \equiv -\bar{m}_a m_b \nabla^b \ell^a, \qquad \text{(expansion and twist of the } \ell \text{ congruence)} \quad (14.9)$$

$$\sigma \equiv -m_a m_b \nabla^b \ell^a, \qquad \text{(shear of the } \ell \text{ congruence)} \quad (14.10)$$

$$\kappa \equiv -m_a \ell_b \nabla^b \ell^a, \qquad \text{(failure of } \ell \text{ to be geodesic)} \quad (14.11)$$

with analogous quantities (μ, λ, ν) controlling the optical properties of n^a.


Directional
derivatives: $D = \ell \cdot \nabla$, $\Delta = n \cdot \nabla$, $\delta = m \cdot \nabla$, $\bar{\delta} = \bar{m} \cdot \nabla$
Spin coefficients encode $\nabla\ell$, ∇n, ∇m:ρ expansion and twist of ℓ congruence;
σ shear of ℓ; κ non-geodesicity of ℓ.Similarly μ, λ, ν for n. Remaining
coefficients $(\epsilon, \gamma, \alpha, \beta, \tau, \pi)$ encode frame acceleration and angular transport.

Conceptual diagram: NP operators and the meaning of spin coefficients as optical data.

Figure 14.3: Conceptual diagram: NP operators and the interpretation of spin coefficients as geometric optics data.

14.5 Two-spinors as square roots of null geometry

14.5.1 Spin spaces and index structure

In four-dimensional Lorentzian geometry, spinors provide a language where vectors can be built from bilinears of spinors. The complex two-dimensional spin space S carries an antisymmetric spinor ϵ_{AB} that raises and lowers indices:

$$\xi_A = \epsilon_{AB}\,\xi^B, \qquad \xi^A = \epsilon^{AB}\,\xi_B, \qquad \epsilon_{AB}\epsilon^{BC} = \delta_A{}^C. \tag{14.12}$$

Complex conjugation produces a separate (primed) spin space with indices A'. The soldering form $\sigma^a{}_{AA'}$ maps spinors to vectors:

$$v^a = \sigma^a{}_{AA'}\,\xi^A\bar{\xi}^{A'}. \tag{14.13}$$

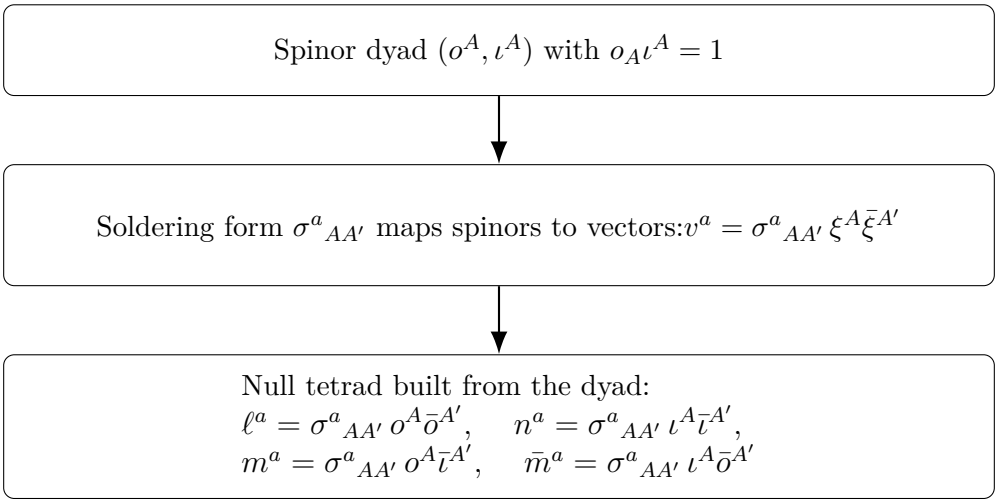
The right-hand side is null when built from a single spinor ξ^A. This is the square-root nature of null directions.

14.5.2 Spin frames and the null tetrad correspondence

Choose a spinor dyad (o^A, ι^A) with normalization $o_A\iota^A = 1$. Then the associated null tetrad is

$$\ell^a = \sigma^a{}_{AA'}\,o^A\bar{o}^{A'}, \qquad n^a = \sigma^a{}_{AA'}\,\iota^A\bar{\iota}^{A'}, \qquad m^a = \sigma^a{}_{AA'}\,o^A\bar{\iota}^{A'}. \tag{14.14}$$

In Kerr, repeated principal null directions align with a preferred spin frame. This alignment is what reduces the Weyl scalars to a single nonzero component in a principal tetrad (type D) [1–3].


Spinor dyad (o^A, ι^A) with $o_A\iota^A = 1$
Soldering form $\sigma^a{}_{AA'}$ maps spinors to vectors:$v^a = \sigma^a{}_{AA'}\,\xi^A\bar{\xi}^{A'}$
Null tetrad built from the dyad:
$\ell^a = \sigma^a{}_{AA'}\,o^A\bar{o}^{A'}, \quad n^a = \sigma^a{}_{AA'}\,\iota^A\bar{\iota}^{A'},$
$m^a = \sigma^a{}_{AA'}\,o^A\bar{\iota}^{A'}, \quad \bar{m}^a = \sigma^a{}_{AA'}\,\iota^A\bar{o}^{A'}$

Conceptual diagram: spinors generate null vectors and encode null geometry.

Figure 14.4: Conceptual diagram: spinor dyad, soldering form, and induced null tetrad.

14.6 Weyl scalars as curvature multipole moments

14.6.1 Vacuum curvature and the Weyl tensor

In vacuum, the Ricci tensor vanishes, so the full curvature content is captured by the Weyl tensor C_{abcd}. NP decomposes C_{abcd} into five complex scalars:

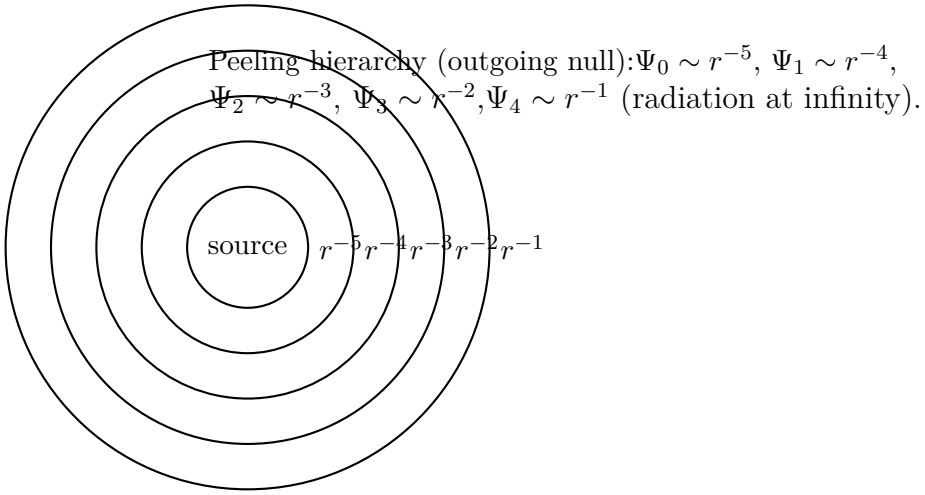
$$\Psi_0 = -C_{abcd}\,\ell^a m^b \ell^c m^d, \tag{14.15}$$
$$\Psi_1 = -C_{abcd}\,\ell^a n^b \ell^c m^d, \tag{14.16}$$
$$\Psi_2 = -C_{abcd}\,\ell^a m^b \bar{m}^c n^d, \tag{14.17}$$
$$\Psi_3 = -C_{abcd}\,\ell^a n^b \bar{m}^c n^d, \tag{14.18}$$
$$\Psi_4 = -C_{abcd}\,n^a \bar{m}^b n^c \bar{m}^d. \tag{14.19}$$

Heuristically, Ψ_2 is Coulomb-like tidal curvature, while Ψ_0 and Ψ_4 encode transverse radiation along ℓ and n. This is sharpened in asymptotically flat spacetimes via peeling [1, 2].

14.6.2 From Ψ_4 to what detectors measure

At future null infinity, a standard relation links the radiative Weyl scalar to strain:

$$\Psi_4 \approx \ddot{h}_+ - \mathrm{i}\,\ddot{h}_\times, \tag{14.20}$$


Peeling hierarchy (outgoing null):$\Psi_0 \sim r^{-5}$, $\Psi_1 \sim r^{-4}$,
$\Psi_2 \sim r^{-3}$, $\Psi_3 \sim r^{-2}$,$\Psi_4 \sim r^{-1}$ (radiation at infinity).
source
$r^{-5}r^{-4}r^{-3}r^{-2}r^{-1}$

Conceptual diagram: peeling suggests why Ψ_4 dominates far from the source.

Figure 14.5: Conceptual diagram: peeling picture of Weyl scalars and asymptotic radiation.

where h_+ and $h_\times$ are the strain polarizations and dots are time derivatives along an appropriate asymptotic time coordinate [1, 4]. The conceptual ordering matters: gravitational radiation is naturally a curvature observable first, and a displacement observable only after integrating in time with boundary conditions. This is one reason tetrads matter: Ψ_4 is a tetrad component of curvature, so its physical meaning depends on the tetrad choice, especially on alignment with outgoing radiation.

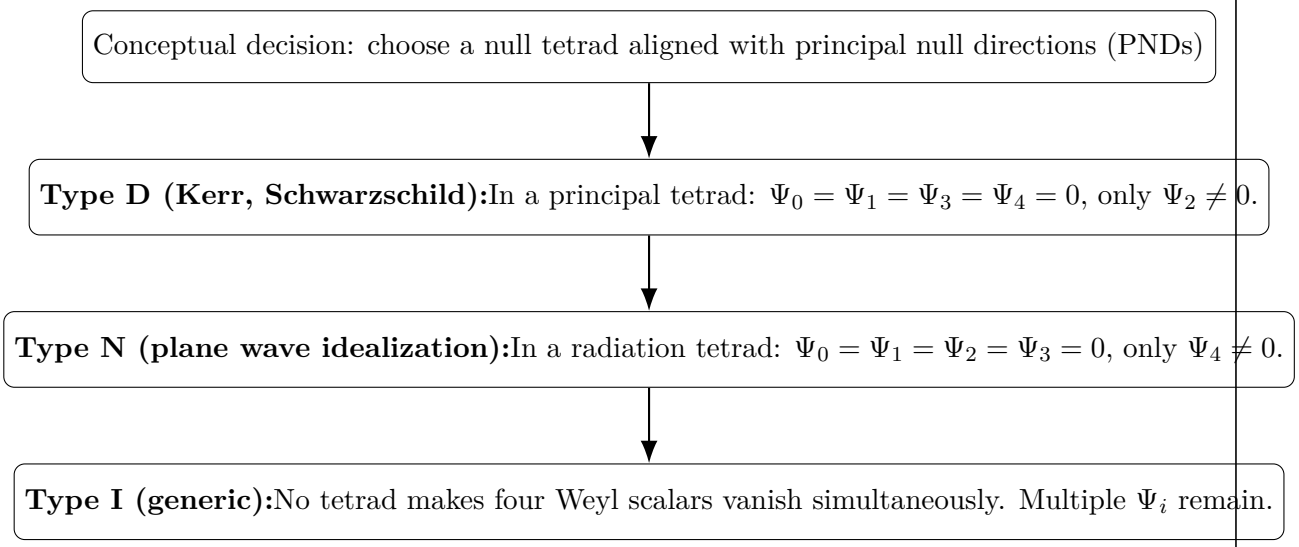
14.7 Petrov classification and Kerr specialness

14.7.1 Algebraic types as statements about principal null directions

Petrov classification organizes Weyl tensors by principal null direction multiplicities. In NP language, this often appears as statements about whether a tetrad exists in which certain Weyl scalars vanish. Type N spacetimes can be framed so only Ψ_4 is nonzero (pure outgoing radiation idealization). Type D spacetimes can be framed so only Ψ_2 is nonzero (Coulomb-like field with repeated principal null directions). Kerr is the archetypal type D spacetime [2, 3].

14.7.2 Why Kerr specialness becomes calculational power

Kerr has two repeated principal null directions. When the null tetrad is aligned with them (for example the Kinnersley tetrad), curvature simplifies: the Weyl scalars collapse


Conceptual decision: choose a null tetrad aligned with principal null directions (PNDs)
Type D (Kerr, Schwarzschild):In a principal tetrad: $\Psi_0 = \Psi_1 = \Psi_3 = \Psi_4 = 0$, only $\Psi_2 \neq 0$.
Type N (plane wave idealization):In a radiation tetrad: $\Psi_0 = \Psi_1 = \Psi_2 = \Psi_3 = 0$, only $\Psi_4 \neq 0$.
Type I (generic):No tetrad makes four Weyl scalars vanish simultaneously. Multiple Ψ_i remain.

Conceptual diagram: Petrov types as statements about which Ψ_i vanish in a special tetrad.

Figure 14.6: Conceptual diagram: Petrov classification intuition using NP Weyl scalars.

to the single Coulomb term Ψ_2. This is not a trick; it is a geometric statement. The payoff is that certain perturbation variables (notably Ψ_0 and Ψ_4) satisfy decoupled, separable equations on Kerr, leading to the Teukolsky master equation [3,4]. This is the bridge from geometry to quasi-normal modes: ringdown is curvature music, and NP is a way to write the instrument.

14.8 Data-based illustrations: catalog-level curvature and radiation scales

Here we use open catalogs as a proxy for curvature scale. The catalogs do not provide $\Psi_i(t)$ values, but they provide what sets the scale of near-horizon curvature and ringdown: mass, spin combinations, redshift, and distance. The dataset `gwtc3_confident_v2.csv` is constructed from the GWOSC GWTC-3 confident event table medians [6]. We keep uncertainties implicit to preserve reproducibility from the public table.

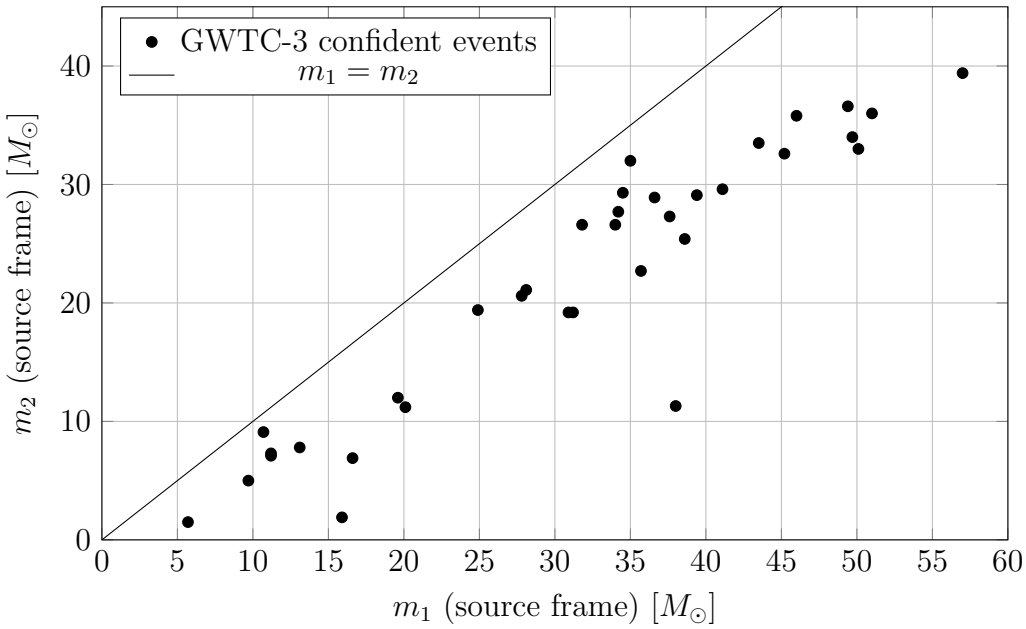


Figure 14.7: Data-based figure: source-frame component masses for GWTC-3 confident events. Values are medians from the GWOSC table [6].

14.8.1 Mass ratios and binary asymmetry in the data

14.8.2 Spin information and why χ_{eff} is a geometric compression

The effective spin parameter χ_{eff} compresses spin information into a single number that strongly influences inspiral phasing and therefore the curvature evolution along the orbit. Even without explicitly constructing a tetrad, χ_{eff} is a proxy for how strongly spin-orbit geometry influences the orbital dynamics. The next plot shows χ_{eff} versus chirp mass $\mathcal{M}$.

14.8.3 Distance, signal-to-noise, and detector tetrads

Detectors measure strain in a laboratory frame, which is a tetrad in practice: the arms define spatial directions and laboratory time defines a timelike direction. The catalog reports network SNR and luminosity distance. The spread at fixed distance is real and encodes orientation and polarization effects.

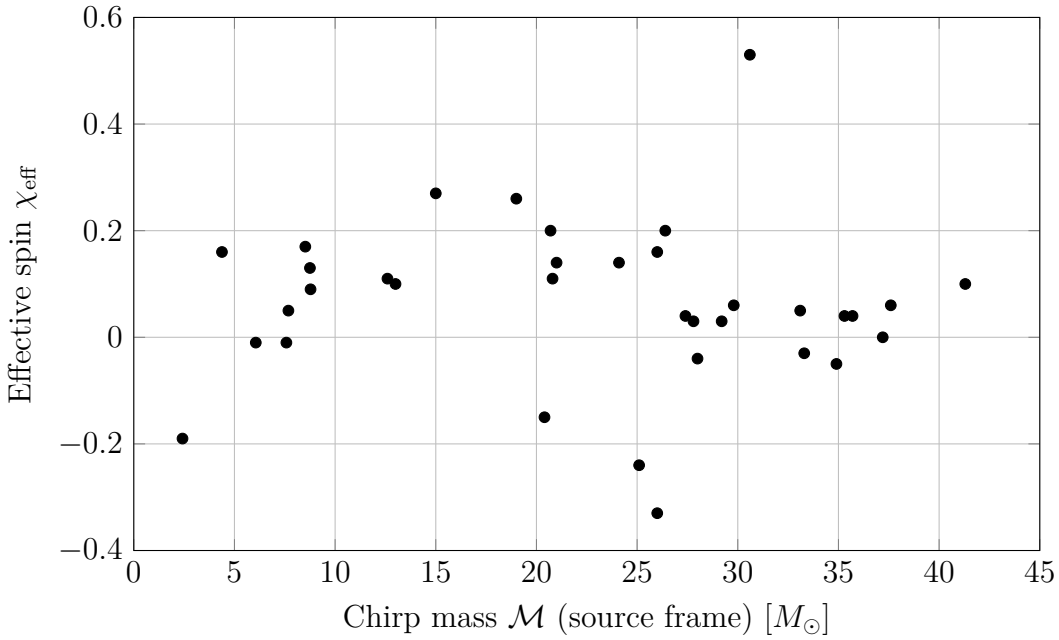


Figure 14.8: Data-based figure: effective spin versus chirp mass for GWTC-3 confident events (medians from GWOSC) [6].

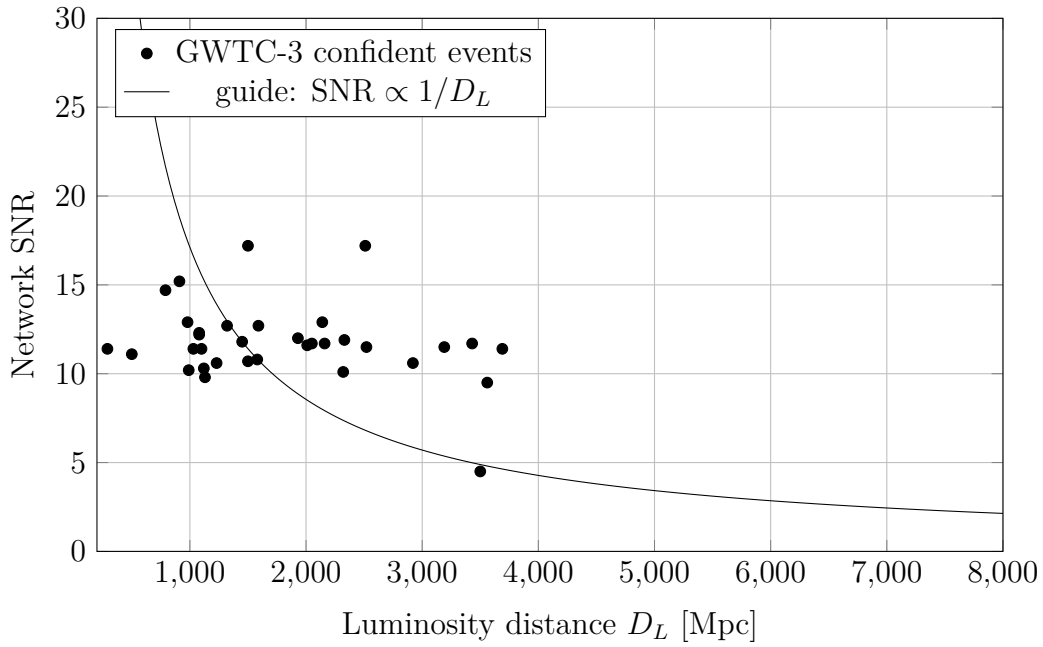


Figure 14.9: Data-based figure: network SNR versus luminosity distance for GWTC-3 confident events [6]. The curve is a guide-to-eye inverse scaling, not a fit.

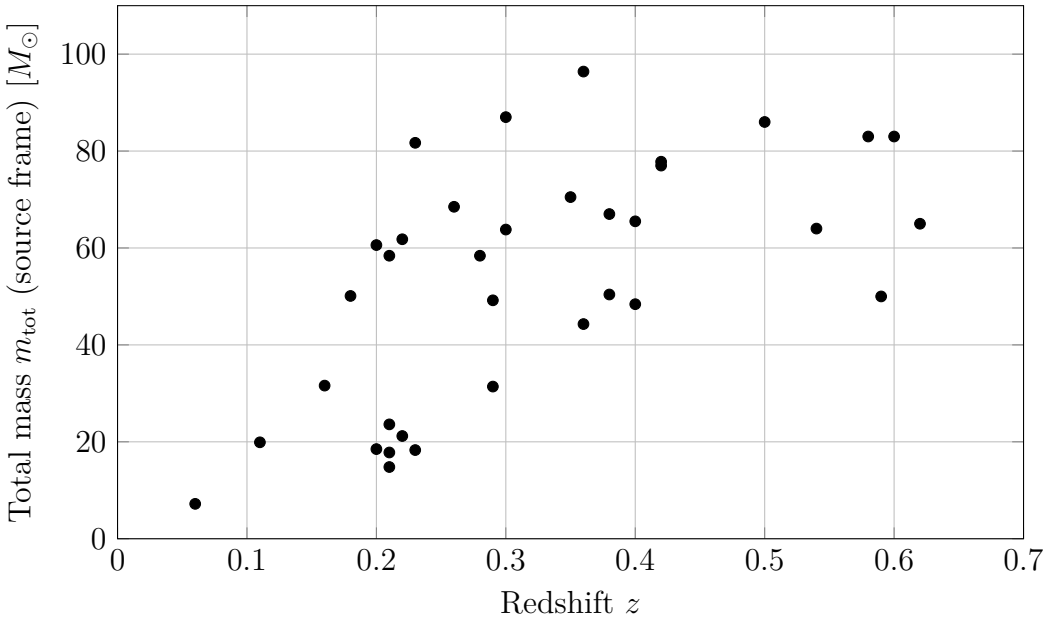


Figure 14.10: Data-based figure: total mass versus redshift for GWTC-3 confident events [6].

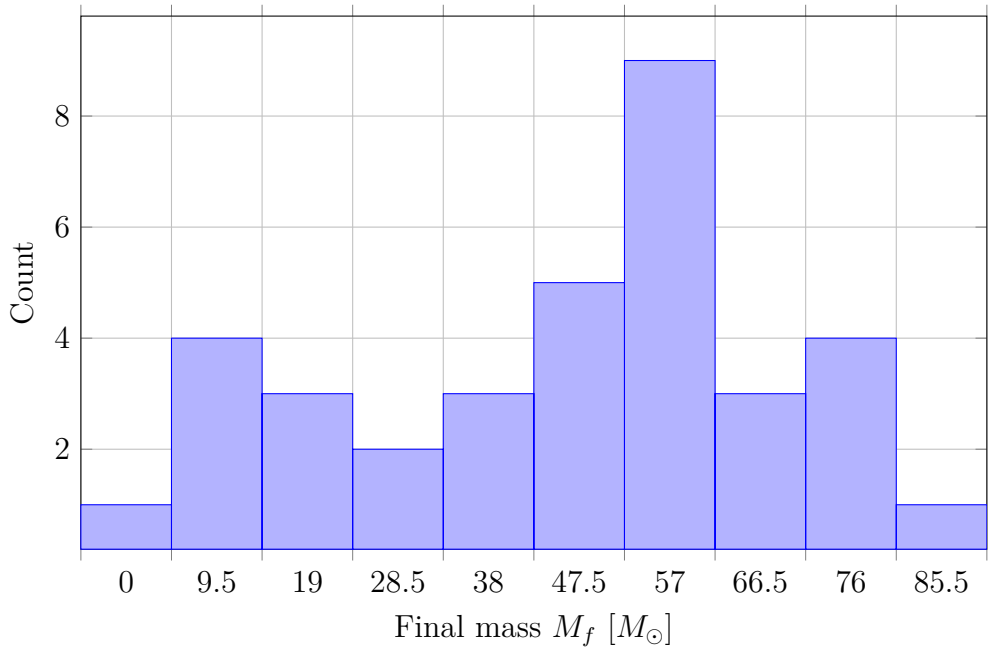


Figure 14.11: Data-based figure: histogram of final (remnant) masses in the GWTC-3 confident sample (medians from GWOSC) [6].

14.8.4 Redshift and total mass

14.8.5 Distribution of remnant masses

14.8.6 Data-derived ringdown scale from remnant mass

For a Schwarzschild black hole, the fundamental quadrupolar quasi-normal mode has a dimensionless frequency $\omega M \approx 0.37367$ (in $G = c = 1$ units), giving the scale estimate

$$f_{220}^{(\mathrm{Schw})} \approx \frac{0.37367}{2\pi} \frac{c^3}{G\,M_f} \approx \frac{12074\ \mathrm{Hz}}{M_f/}. \tag{14.21}$$

This is a scale indicator, not a Kerr fit. We use it to map catalog remnant masses into an observationally anchored frequency scale [5].

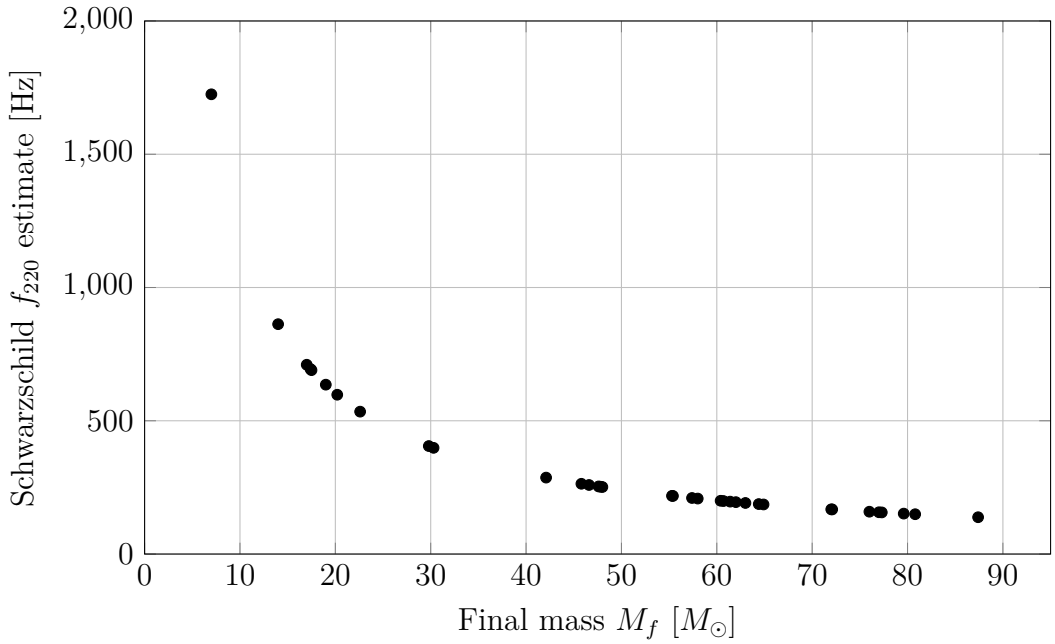


Figure 14.12: Data-based figure: ringdown frequency scale computed from GWTC-3 remnant masses using Eq. (14.21). Data: GWOSC [6]. Constant source: standard QNM results summarized in [5].

14.9 Applications and phenomenon statements grounded in real data

14.9.1 Phenomenon 1: detector polarizations as tetrad projections

Phenomenon statement: Strain polarizations arise from projecting curvature onto a detector-adapted tetrad

A ground-based interferometer defines a preferred local tetrad: the timelike leg is laboratory time and two spatial legs are arm directions. The detector output is not the full spacetime curvature. It is a specific projection integrated into strain. The relation $\Psi_4 \approx \ddot{h}_+ - i\ddot{h}_\times$ expresses a hierarchy: curvature is primary, strain is secondary, and polarization is a statement about decomposition relative to the transverse complex basis $(m, \bar{m})$. In practice this means the same astrophysical event can appear stronger or weaker depending on polarization alignment with the detector network.

The GWTC-3 catalog illustrates this geometry in aggregate. Figure 14.9 shows that distance alone does not determine SNR. The spread is a real-world imprint of orientation and polarization. In tetrad language, the variation is expected: different source orientations correspond to different radiative curvature components presented to the detector basis. NP language does not remove the complexity, but it clarifies what the complexity is: it is the relative geometry of two tetrads, one tied to the source radiation frame and one tied to the detector.

14.9.2 Phenomenon 2: ringdown spectroscopy as curvature eigenmodes

Phenomenon statement: Ringdown frequencies encode remnant geometry and are naturally written in NP variables

The ringdown of a black hole is a relaxation of curvature. It is not a vibration of matter, but an eigenmode spectrum of the spacetime geometry. In Kerr, the Teukolsky equation provides a decoupled description of curvature perturbations in terms of Ψ_0 or Ψ_4 [3, 4]. Solutions decompose into quasi-normal modes whose complex frequencies depend on remnant mass and spin. This dependence allows detectors to infer final parameters from late-time signal structure and, in principle, to test the Kerr hypothesis by measuring multiple modes.

Our data-based plots show the scale separation at the catalog level. Figure 14.11 shows remnant masses from a few to nearly 90 , implying a wide range of curvature time scales. Figure 14.12 turns those masses into a ringdown frequency scale using a standard constant [5]. The inverse relationship between mass and frequency is not a modeling artifact; it is dimensional. In $G = c = 1$ units, a black hole has only one length scale, so every curvature oscillation frequency scales like $1/M$.

14.9.3 Phenomenon 3: stellar-mass black holes in binaries as horizon laboratories

X-ray binaries provide long-lived laboratories where horizon-scale geometry coexists with rich astrophysical environments. The stellarcollapse.org compilation provides a publicly accessible table and PDF of observed black hole masses in X-ray binaries [8]. Many entries are ranges or bounds. To produce plot-ready material while preserving scale information, we apply a transparent conversion rule: midpoints for ranges and quoted bounds as central values for bounded entries. This is not claiming new measurements; it is transforming a public table into a consistent numeric dataset for visualization.

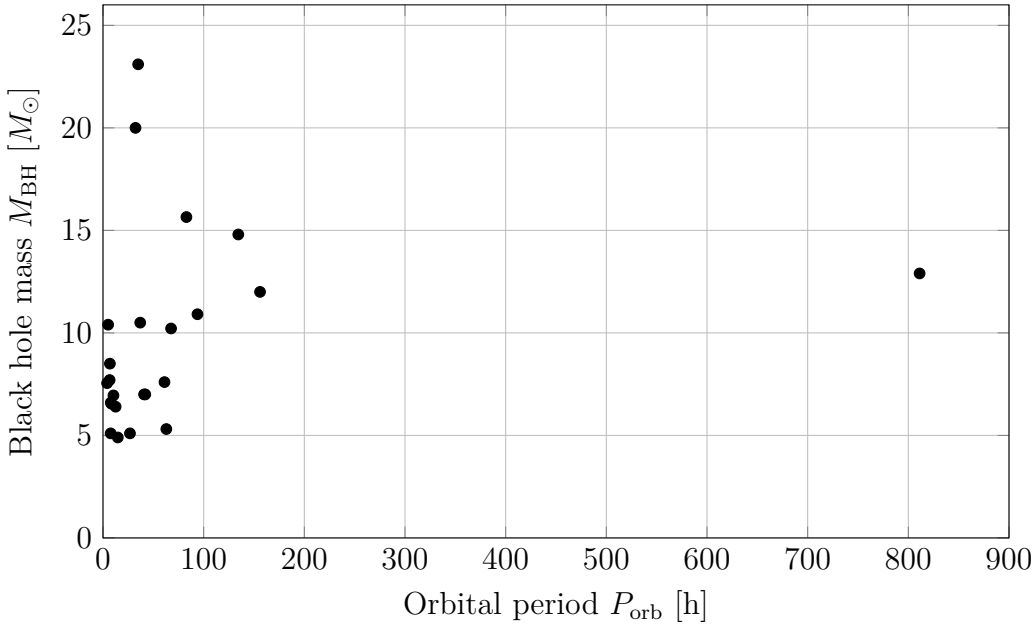


Figure 14.13: Data-based figure: black hole mass versus orbital period for a compiled X-ray binary sample. Data source: stellarcollapse.org observed black hole masses [8].

Phenomenon statement: Mass measurements are curvature scale measurements once the radius scale is fixed

A black hole of mass M sets a horizon length scale $r_s = 2GM/c^2$ and a curvature time scale $t_M = GM/c^3$. These scalings explain why stellar-mass black holes produce millisecond variability while supermassive black holes produce variability on hours to days. In NP terms, the Coulomb-like curvature scalar Ψ_2 for a stationary black hole scales roughly like M/r^3. A mass measurement therefore becomes a curvature scale measurement as soon as one decides which radius scale is operationally relevant, for example near the horizon, near the innermost stable orbit, or in a radiative extraction zone.

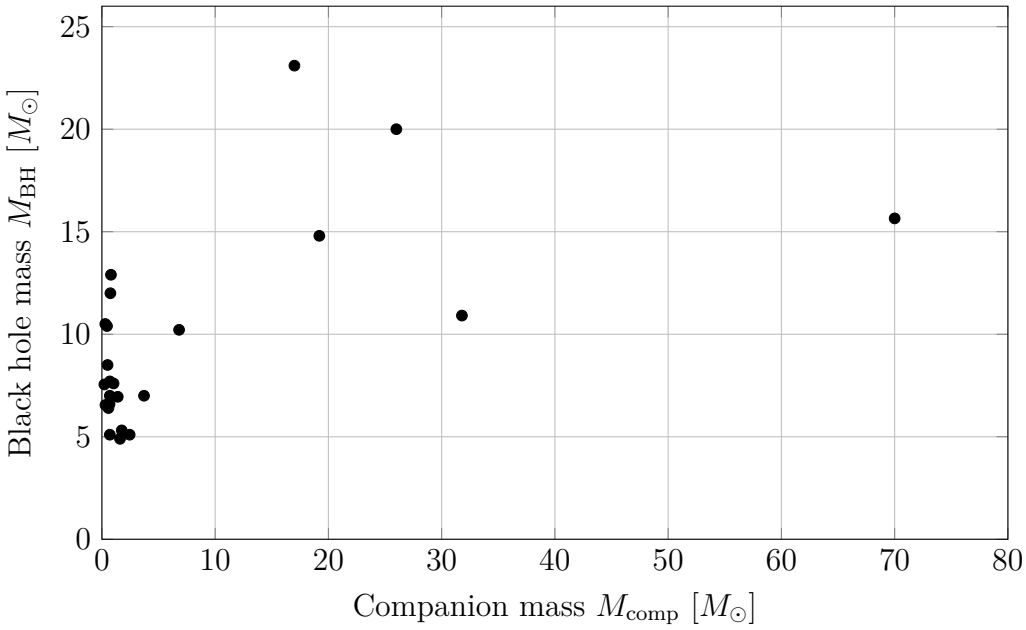


Figure 14.14: Data-based figure: companion mass versus black hole mass for the same X-ray binary compilation [8].

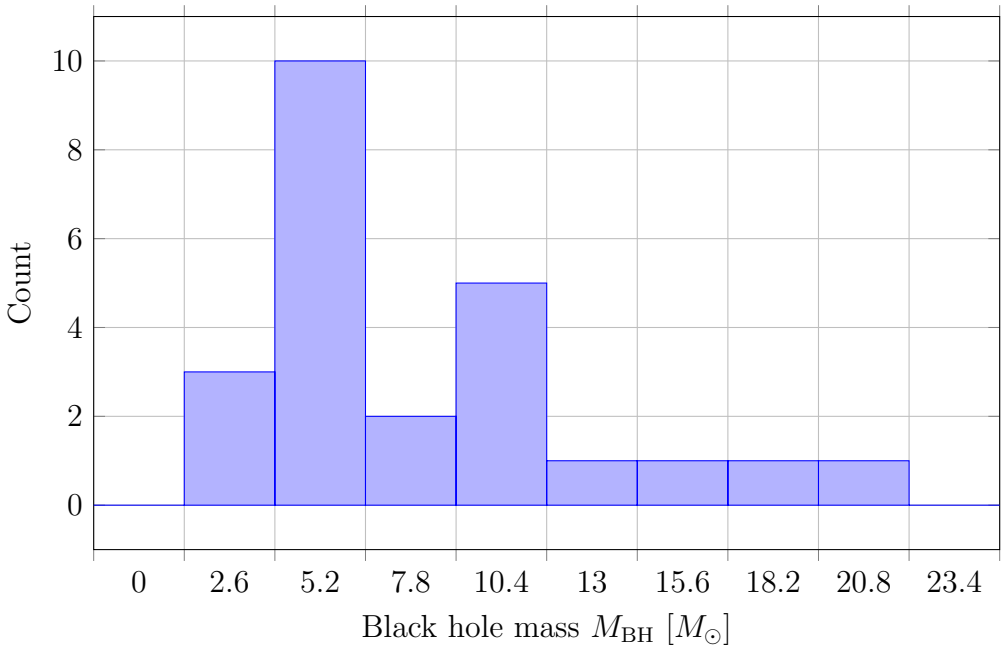


Figure 14.15: Data-based figure: histogram of stellar-mass black hole masses in the compiled X-ray binary sample [8].

14.9.4 Phenomenon 4: horizon-scale imaging and the universality of the ring diameter

Phenomenon statement: EHT ring diameters compare across orders of magnitude because the geometry is dimensionless

A striking observational fact is that horizon-scale images can be compared across many orders of magnitude in mass. The reason is geometric: the angular scale is set by

$$\theta_g \equiv \frac{GM}{Dc^2}, \tag{14.22}$$

and the Schwarzschild shadow diameter is $d_{\rm sh} = 6\sqrt{3}\,\theta_g$. Spin modifies the diameter weakly compared to its dependence on θ_g.

The EHT measured an emission ring diameter of about 42 ± 3 for M87 and about 51.8 ± 2.3 for Sagittarius A [9, 10]. Those values are not merely image facts; they are horizon-geometry facts. They assert that the observed ring sizes are consistent with general-relativistic photon orbit structure when independently measured mass and distance are inserted into θ_g.

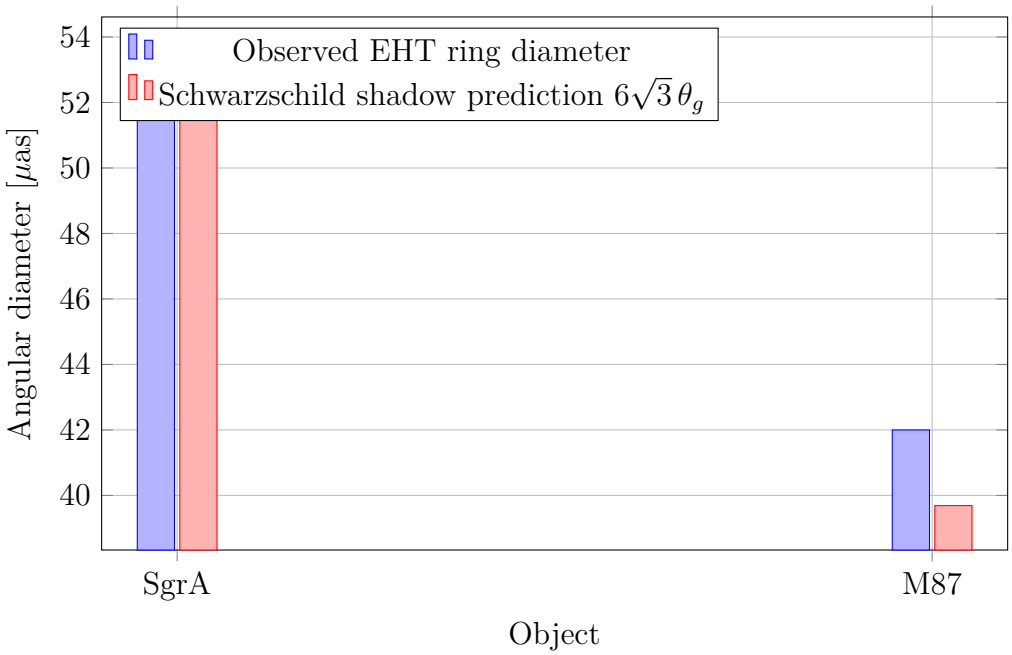


Figure 14.16: Data-based figure: EHT observed ring diameters compared to a Schwarzschild shadow diameter prediction computed from masses and distances quoted in EHT publications [9, 10].

14.9.5 Phenomenon 5: why NP curvature scalars are useful in numerical and data pipelines

Phenomenon statement: Wave extraction prefers curvature first, strain second

In numerical relativity, a common workflow extracts gravitational radiation as Ψ_4 on large-radius spheres and then integrates to obtain strain. This is not only a numerical convenience; it is geometrically natural. Ψ_4 is a component of the Weyl tensor in an outgoing radiation tetrad, so it isolates radiative curvature in a way that is invariant under many gauge distortions that complicate metric-based extraction. The tetrad choice still matters, but NP language makes the dependence explicit and controllable. Once Ψ_4 is available, the strain can be produced by integration with physically motivated boundary conditions. This matches the conceptual hierarchy of Eq. (14.20) and aligns naturally with detector interpretation, where strain is a derived observable.

14.10 Conclusion

Tetrads are the geometry of measurement. Null tetrads are the geometry of horizons and radiation. Spinors are square roots of null directions, and Newman-Penrose is the operational calculus binding these statements into computable curvature components. In this chapter we built the language while keeping physical meaning visible: connection coefficients become spin coefficients with optical interpretations, and curvature becomes Weyl scalars with radiative interpretations.

The deeper message is that NP is not merely a notational convenience. It is a way to make Einstein geometry speak the language of observation. Gravitational-wave catalogs provide masses, spin combinations, redshifts, and distances. Those are precisely the inputs that set curvature scales and ringdown time scales. X-ray binaries anchor long-lived stellar-mass horizons in astrophysical environments. Horizon-scale imaging demonstrates a universality: across vastly different masses, observed ring diameters are consistent with dimensionless shadow predictions expressed through $\theta_g = GM/(Dc^2)$. These connections make the NP formalism feel less like an abstract calculus and more like a geometric interface between theory, simulation, and data.

Manual index of key terms (by section)

Term	Where discussed
Tetrad, vierbein, orthonormal frame	Sections 14.3-14.3.2; Conceptual Fig. 14.1

Coframe, Cartan structure equations	Section 14.3.2; Eqs. (14.4)-(14.5)
Null tetrad $(\ell, n, m, \bar{m})$	Sections 14.4-14.4.2; Conceptual Fig. 14.2
Spin coefficients	Section 14.4.2; Conceptual Fig. 14.3
Two-spinors, soldering form	Sections 14.5-14.5.2; Conceptual Fig. 14.4
Weyl scalars Ψ_0 to Ψ_4	Sections 14.6-14.6.2; Conceptual Fig. 14.5
Petrov classification	Section 14.7; Conceptual Fig. 14.6
Ringdown scale estimate	Section 14.8; Eq. (14.21); Data Fig. 14.12
EHT ring diameter and shadow prediction	Section 14.9; Data Fig. 14.16

Bibliography

[1] E. Newman and R. Penrose, "An approach to gravitational radiation by a method of spin coefficients," *Journal of Mathematical Physics*, vol. 3, no. 3, pp. 566–578, 1962. [Online]. Available: https://pubs.aip.org/aip/jmp/article-abstract/3/3/566/387969/An-Approach-to-Gravitational-Radiation-by-a-Method

[2] R. Penrose and W. Rindler, *Spinors and Space-Time, Volume 1: Two-Spinor Calculus and Relativistic Fields*. Cambridge University Press, 1984. [Online]. Available: https://www.amazon.com/Spinors-Space-Time-Two-Spinor-Relativistic-Mathematical/dp/0521245273

[3] S. Chandrasekhar, *The Mathematical Theory of Black Holes*. Oxford University Press, 1983. [Online]. Available: https://global.oup.com/academic/product/the-mathematical-theory-of-black-holes-9780198503705

[4] S. A. Teukolsky, "Perturbations of a rotating black hole. I. Fundamental equations for gravitational, electromagnetic, and neutrino-field perturbations," *The Astrophysical Journal*, vol. 185, pp. 635–647, 1973. [Online]. Available: https://ui.adsabs.harvard.edu/abs/1973ApJ...185..635T/abstract

[5] E. Berti, V. Cardoso, and A. O. Starinets, "Quasinormal modes of black holes and black branes," *Classical and Quantum Gravity*, vol. 26, 163001, 2009. [Online]. Available: https://doi.org/10.1088/0264-9381/26/16/163001 and https://arxiv.org/abs/0905.2975

[6] Gravitational Wave Open Science Center (GWOSC), "GWTC-3-confident event table (public catalog parameters)," Event Portal. [Online]. Available: https://gwosc.org/eventapi/html/GWTC-3-confident/.

[7] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run," *Physical Review X*, vol. 13, 041039, 2023. [Online] Available: https://arpi.unipi.it/retrieve/2df7b488-b663-48e9-a611-82d91cbac428/PhysRevX.13.041039.pdf

[8] stellarcollapse.org, "Observed Black Hole Masses (X-ray binaries) table and PDF," compilation page. [Online]. Available: https://stellarcollapse.org/bhmasses.html.

[9] Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole," *The Astrophysical Journal Letters*, vol. 875, L1, 2019. [Online]. Available: https://arxiv.org/abs/1906.11238

[10] Event Horizon Telescope Collaboration, "First Sagittarius A* Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole in the Center of the Milky Way," 2023. [Online]. Available: https://arxiv.org/abs/2311.08680 and https://eventhorizontelescope.org/publications/first-sagittarius-event-horizon-telescope-results-i-shadow-supermassive-black-hole

Index

Cartan equations, 20

EHT, 20

null tetrad, 20

Petrov classification, 20

ringdown, 20

shadow diameter, 20
soldering form, 20
spin coefficients, 20
spinor, 20

tetrad, 19

vierbein, 19

Weyl scalars, 20

Contents

15 Perturbations of Black Hole Geometry **1**

15.1 Introduction: ripples as geometry with a memory 1

15.2 Linearized Einstein equations and gauge as geometry 2

15.2.1 The linearization map and why it is not coordinate calculus . . . 2

15.2.2 Gauge transformations as infinitesimal diffeomorphisms 3

15.3 Schwarzschild perturbations: Regge-Wheeler and Zerilli as curvature scattering 4

15.3.1 Parity decomposition and master variables 4

15.3.2 Geometric meaning of the potentials and the photon sphere . . . 5

15.4 Kerr perturbations: Newman-Penrose curvature and the Teukolsky equation 6

15.4.1 Why rotating black holes demand a different viewpoint 6

15.4.2 The Teukolsky master equation as a statement of separability . . 7

15.5 Quasinormal modes: spectral geometry of horizons 8

15.5.1 Boundary conditions that create a discrete spectrum 8

15.5.2 A practical fitting formula for the dominant Kerr mode 8

15.6 Data-driven illustrations from GWTC-1: from catalog values to ringdown scales 9

15.6.1 Detection strength and observing time 11

15.6.2 Mass scale: chirp mass distribution 11

15.6.3 Spin information in the inspiral: effective spin versus mass ratio . 11

15.6.4 Localization geometry: sky area versus SNR 13

15.6.5 Energy budget: radiated energy fraction 13

15.6.6 Peak luminosity scale: curvature radiation as a burst 13
15.6.7 Mass-energy closure check using catalog medians 13
15.6.8 Ringdown frequency scale from catalog remnant parameters . . . 13
15.6.9 Cosmological redshift as a geometric rescaling of ringdown 13
15.6.10 Damping and quality: how spin controls decay 13
15.7 Stability: what "small ripples do not destroy the hole" really means . . . 17
15.7.1 Mode stability, boundedness, decay, and nonlinear stability 17
15.7.2 Interpretation through horizons and causal structure 17
15.8 Conclusion 18

Chapter 15

Perturbations of Black Hole Geometry

15.1 Introduction: ripples as geometry with a memory

A black hole spacetime is not merely a solution of Einstein's equations that happens to be pretty. It is a geometric object with a causal boundary, a set of curvature invariants, and a collection of preferred structures determined by symmetry and by the presence of a horizon. Yet nature rarely hands us an eternal, exactly stationary metric. Real black holes are formed by collapse, disturbed by infalling matter, torqued by companions, and kicked by asymmetric radiation. The mathematical language that lets us stay close to exact geometry while still describing these real disturbances is *perturbation theory*. In its cleanest form, perturbation theory begins with a background metric $g^{(0)}_{ab}$, chosen because it captures the dominant geometric features of the system, and then adds a small correction h_{ab} so that

$$g_{ab} = g^{(0)}_{ab} + \epsilon\, h_{ab} + \mathcal{O}(\epsilon^2), \tag{15.1}$$

where ϵ is a book-keeping parameter that marks order in smallness. The physical meaning of ϵ depends on context. It might represent the mass ratio in an extreme mass ratio inspiral, the amplitude of tidal distortion from a companion, or the post-merger deviation of a newly formed remnant from its final Kerr state.

The core conceptual point is that h_{ab} is not a collection of coordinate components, but a tensor field on the same manifold. If we push the background geometry forward by an infinitesimal diffeomorphism, we change the components of h_{ab} even when the physical perturbed spacetime is the same. This is gauge freedom, and in perturbation theory it is not optional. It is the mathematical statement that coordinates remain a polite lie even when the deviation from the truth is small.

The pay-off for taking this geometric view is enormous. In Schwarzschild, spherical

symmetry lets us decompose perturbations into tensor spherical harmonics and separate the dynamics into two parity sectors governed by one-dimensional wave equations with curvature potentials, the Regge-Wheeler and Zerilli equations [3, 4, 7]. In Kerr, the correct objects to perturb are not metric components but curvature scalars in a null tetrad, leading to the Teukolsky master equation [5, 6]. In both cases, the response of the geometry has a characteristic spectral signature: quasinormal modes, damped oscillations whose complex frequencies are determined by the horizon boundary condition and the asymptotics at infinity [9, 11].

Phenomenon statement 15.1: Ringdown as curvature music in detectors

When LIGO and Virgo measure a post-merger gravitational-wave signal, the late-time waveform is well described by a sum of damped sinusoids. Those sinusoids are not arbitrary. They are the eigenfrequencies of linear perturbations of the final Kerr geometry. In other words, a detector measures a spectral property of spacetime curvature, not a mechanical vibration of a material object. This is why the ringdown is often called black hole spectroscopy [10, 15].

This chapter is written with two goals. The first is conceptual: to show how linearized gravity becomes a geometry of gauge orbits, master variables, and boundary conditions. The second is empirical: to ground the discussion in open observational data. We therefore include ten data-based figures generated directly in LaTeX from a dataset extracted from the GWTC-1 binary black hole events hosted by the Gravitational Wave Open Science Center [1, 2]. Several derived quantities, such as predicted ringdown frequencies, are computed from the data using published fitting formulas for Kerr quasinormal modes [10]. The purpose is not to replace full parameter estimation, but to show that the geometric ideas of perturbations map onto measurable patterns in real events.

15.2 Linearized Einstein equations and gauge as geometry

15.2.1 The linearization map and why it is not coordinate calculus

Let $G_{ab}[g]$ denote the Einstein tensor constructed from a metric g_{ab}. The Einstein equations read

$$G_{ab}[g] = 8\pi\, T_{ab}. \tag{15.2}$$

To linearize, we consider $g_{ab}(\epsilon) = g^{(0)}_{ab} + \epsilon h_{ab}$ and expand:

$$G_{ab}[g^{(0)} + \epsilon h] = G_{ab}[g^{(0)}] + \epsilon\, \delta G_{ab}[h] + \mathcal{O}(\epsilon^2). \tag{15.3}$$

If the background is a vacuum black hole solution, then $G_{ab}[g^{(0)}] = 0$ and the linearized vacuum equations are simply

$$\delta G_{ab}[h] = 0. \tag{15.4}$$

Equation (15.4) is often written as a complicated expression in coordinate components, but the geometric content is simpler: it is a linear differential operator on symmetric $(0,2)$ tensor fields. It is constrained by the linearized Bianchi identity $\nabla^a \delta G_{ab} = 0$, which is the shadow of diffeomorphism invariance.

A practical way to reveal the wave-like character of (15.4) is to introduce the trace-reversed perturbation $\bar{h}_{ab} = h_{ab} - \frac{1}{2} g^{(0)}_{ab} h$, where $h = g^{(0)\,ab} h_{ab}$. In a Lorenz gauge, defined by

$$\nabla^a \bar{h}_{ab} = 0, \tag{15.5}$$

the linearized equations on a vacuum background reduce to a tensor wave equation with curvature coupling:

$$\nabla^c \nabla_c \bar{h}_{ab} + 2 R^{(0)}{}_a{}^c{}_b{}^d \, \bar{h}_{cd} = 0. \tag{15.6}$$

Here $R^{(0)}{}_a{}^c{}_b{}^d$ is the Riemann tensor of the background. Equation (15.6) is a direct and vivid statement: even when the perturbation is a wave, the background curvature scatters it and couples its components.

15.2.2 Gauge transformations as infinitesimal diffeomorphisms

The geometric origin of gauge freedom is that the physical spacetime is an equivalence class under diffeomorphisms. If we shift coordinates by an infinitesimal vector field ξ^a, the perturbation transforms as

$$h_{ab} \mapsto h'_{ab} = h_{ab} + \nabla_a \xi_b + \nabla_b \xi_a. \tag{15.7}$$

Equation (15.7) is not a trick but a definition: two perturbations related by (15.7) represent the same physical first-order geometry. The operator δG_{ab} respects this redundancy: $\delta G_{ab}[\nabla_a \xi_b + \nabla_b \xi_a] = 0$ on vacuum backgrounds, reflecting diffeomorphism invariance.

A mature perturbation theory therefore has two intertwined tasks. One task is to choose a gauge that makes the equations solvable, such as Lorenz gauge (15.5) or a Regge-Wheeler gauge in spherical symmetry. The other task is to identify gauge-invariant quantities, so that the final physical statements do not depend on the chosen gauge. The canonical examples are the Regge-Wheeler and Zerilli master functions in Schwarzschild [3,4,7], and the curvature scalars ψ_0 and ψ_4 in Kerr [5].

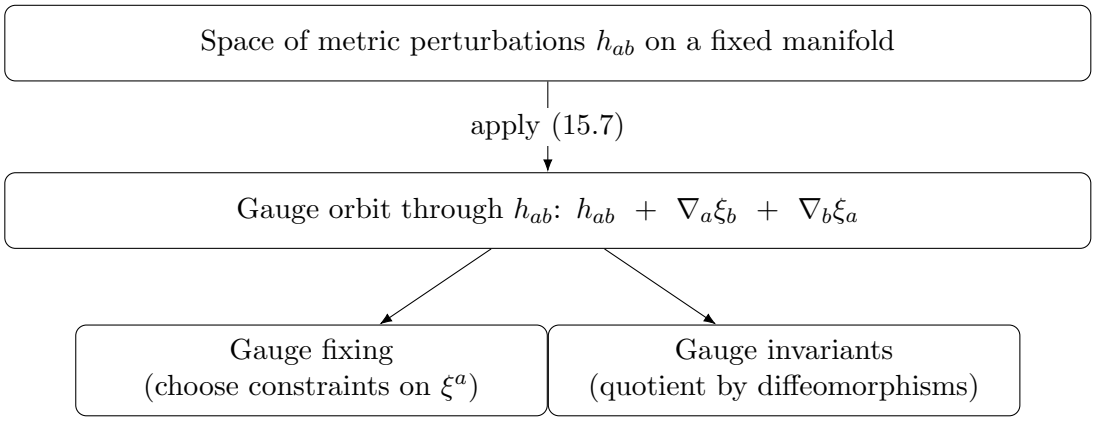

Space of metric perturbations h_{ab} on a fixed manifold
apply (15.7)
Gauge orbit through h_{ab}: $h_{ab} + \nabla_a\xi_b + \nabla_b\xi_a$
Gauge fixing
(choose constraints on ξ^a)
Gauge invariants
(quotient by diffeomorphisms)

Figure 15.1: Conceptual diagram: gauge freedom in linearized gravity as a geometric quotient. Gauge is not a physical ambiguity but a statement that first-order perturbations are defined up to infinitesimal diffeomorphisms.

Phenomenon statement 15.2: Gauge is why numerical relativity waveforms have to be extracted carefully

In a binary black hole simulation, the metric is evolved in a chosen coordinate gauge. The gravitational radiation measured at infinity, however, should not depend on that gauge. This is why waveform extraction is performed using gauge-invariant objects such as the Newman-Penrose scalar ψ_4 or curvature invariants, and why perturbation theory is a natural language to interpret the late-time remnant as a perturbed Kerr geometry [5, 11].

15.3 Schwarzschild perturbations: Regge-Wheeler and Zerilli as curvature scattering

15.3.1 Parity decomposition and master variables

The Schwarzschild metric is spherically symmetric and static. This symmetry is not only a convenience; it is a structural theorem for perturbations. Any tensor field on a spherically symmetric background can be decomposed into tensor spherical harmonics labeled by (ℓ, m), and the perturbations split into two decoupled classes under parity: odd (axial) and even (polar) [3, 4].

Odd parity perturbations change sign under $(\theta, \phi) \mapsto (\pi - \theta, \phi + \pi)$, while even parity perturbations do not. The remarkable fact discovered by Regge and Wheeler is that the entire odd parity sector is governed by a single gauge-invariant master function $\Psi^{(\mathrm{odd})}_{\ell m}(t, r)$

satisfying a one-dimensional wave equation in the tortoise coordinate r_*:

$$\left[-\partial_t^2 + \partial_{r_*}^2 - V_\ell^{(\mathrm{RW})}(r)\right] \Psi_{\ell m}^{(\mathrm{odd})} = S_{\ell m}^{(\mathrm{odd})}. \tag{15.8}$$

A standard explicit form of the Regge-Wheeler potential is

$$V_\ell^{(\mathrm{RW})}(r) = \left(1 - \frac{2M}{r}\right)\left[\frac{\ell(\ell+1)}{r^2} - \frac{6M}{r^3}\right], \tag{15.9}$$

and the tortoise coordinate is defined by

$$r_* = r + 2M \ln\left|\frac{r}{2M} - 1\right|. \tag{15.10}$$

These expressions make the geometric content visible: the same factor $1 - 2M/r$ that defines the horizon also shapes the potential barrier that scatters waves.

Zerilli then showed that the even parity sector also reduces to a single master function $\Psi_{\ell m}^{(\mathrm{even})}$ satisfying a similar equation,

$$\left[-\partial_t^2 + \partial_{r_*}^2 - V_\ell^{(\mathrm{Z})}(r)\right] \Psi_{\ell m}^{(\mathrm{even})} = S_{\ell m}^{(\mathrm{even})}, \tag{15.11}$$

with a different potential $V_\ell^{(\mathrm{Z})}$ [4, 7]. One explicit form is

$$V_\ell^{(\mathrm{Z})}(r) = \left(1 - \frac{2M}{r}\right)\frac{2\lambda^2(\lambda+1)r^3 + 6\lambda^2 M r^2 + 18\lambda M^2 r + 18M^3}{r^3(\lambda r + 3M)^2}, \tag{15.12}$$

where $\lambda = \frac{1}{2}(\ell-1)(\ell+2)$. These are not merely PDEs. They are the statement that Schwarzschild perturbations can be thought of as waves in a one-dimensional effective geometry, scattered by an obstacle made of curvature.

15.3.2 Geometric meaning of the potentials and the photon sphere

The potentials $V_\ell^{(\mathrm{RW})}$ and $V_\ell^{(\mathrm{Z})}$ peak near the photon sphere at $r = 3M$. This is not an accident but a geometric echo: null geodesics near $r = 3M$ are unstable, and wave packets spend extra time there before escaping or falling through the horizon. In the frequency domain, this creates resonant structures that are closely related to quasinormal modes, and in the time domain it produces the characteristic ringdown tail after a disturbance. In a modern gauge-invariant treatment, such as the Martel-Poisson framework, one can construct master functions and fluxes in a way that is explicitly covariant and suitable for practical computation [7].

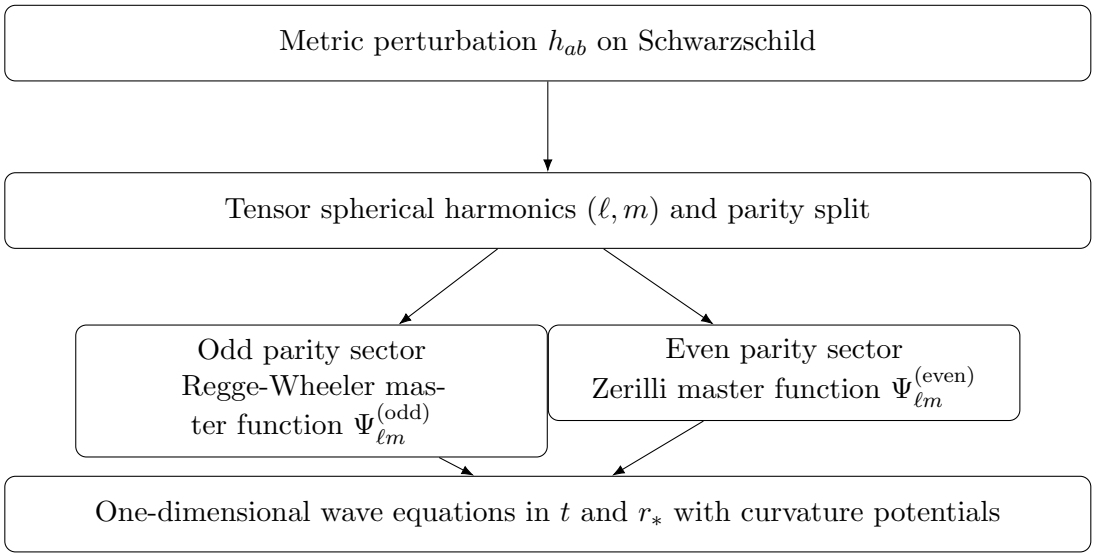

Metric perturbation h_{ab} on Schwarzschild
Tensor spherical harmonics (ℓ, m) and parity split
Odd parity sector
Regge-Wheeler master function $\Psi^{(\mathrm{odd})}_{\ell m}$
Even parity sector
Zerilli master function $\Psi^{(\mathrm{even})}_{\ell m}$
One-dimensional wave equations in t and r_* with curvature potentials

Figure 15.2: Conceptual diagram: the Schwarzschild perturbation pipeline. Symmetry reduces tensorial Einstein perturbations to scalar master equations with curvature potentials.

Phenomenon statement 15.3: Curvature potentials are why ringdown looks universal

If you drop different objects into a Schwarzschild black hole, the details of the initial disturbance differ, but the late-time response tends to a universal form dominated by the same quasinormal spectrum. The reason is geometric: the curvature potential near the photon sphere acts like a filter. Many distinct initial data sets are mapped, through scattering, into a similar outgoing waveform at infinity, dominated by the least damped modes [8, 9].

15.4 Kerr perturbations: Newman-Penrose curvature and the Teukolsky equation

15.4.1 Why rotating black holes demand a different viewpoint

Kerr spacetime is stationary and axisymmetric but not spherically symmetric. The loss of spherical symmetry breaks the clean tensor harmonic decomposition that makes Schwarzschild perturbations so transparent. A naive attempt to perturb metric components in Boyer-Lindquist coordinates leads to a coupled system that does not separate cleanly.

The deep cure is to change the objects we perturb. In the Newman-Penrose formalism, we replace coordinate basis vectors by a null tetrad $(\ell^a, n^a, m^a, \bar{m}^a)$ and project curvature into complex scalars. Of particular relevance for gravitational radiation are the Weyl scalars ψ_0 and ψ_4, which encode, in appropriate regimes, the ingoing and outgoing radiative degrees

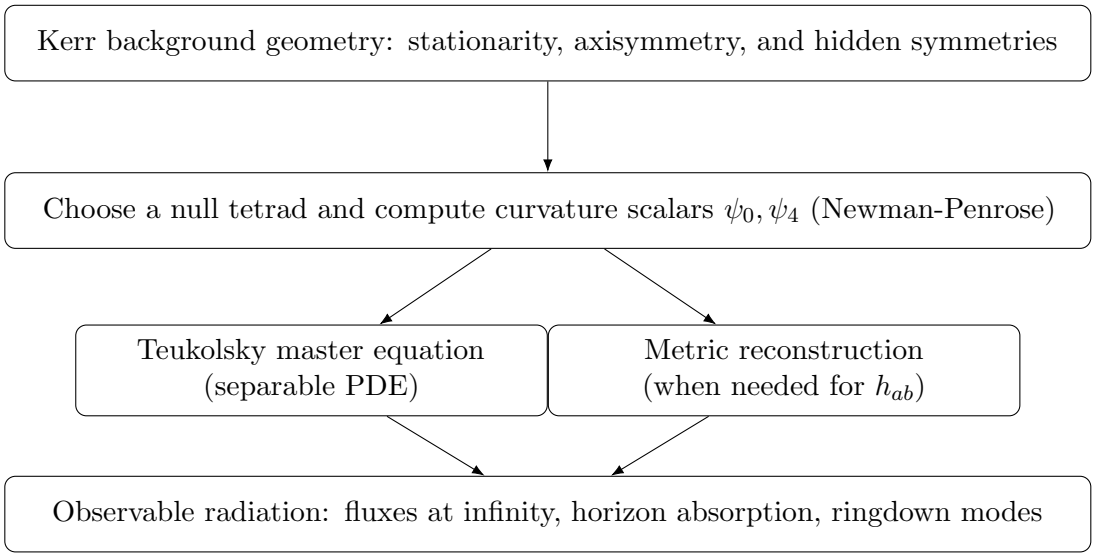

Kerr background geometry: stationarity, axisymmetry, and hidden symmetries
Choose a null tetrad and compute curvature scalars ψ_0, ψ_4 (Newman-Penrose)
Teukolsky master equation (separable PDE)
Metric reconstruction (when needed for h_{ab})
Observable radiation: fluxes at infinity, horizon absorption, ringdown modes

Figure 15.3: Conceptual diagram: why Kerr perturbation theory is naturally curvature-based. The Teukolsky equation packages the radiative degrees of freedom into separable scalars rather than metric components.

of freedom [6]. Teukolsky's key result is that these scalars satisfy a single separable master equation on the Kerr background [5].

15.4.2 The Teukolsky master equation as a statement of separability

The Teukolsky equation is a second-order linear PDE for a field of spin weight s (with $s = -2$ corresponding to ψ_4). Its separability into radial and angular parts is one of the most important algebraic miracles of Kerr geometry and is connected to hidden symmetries and the special Petrov type D structure discussed in Chapter 14. In practice, separability means that one can search for solutions of the form

$$\psi_4 \sim e^{-i\omega t} e^{im\phi} S_{s\ell m}(\theta;\, a\omega)\, R_{s\ell m}(r;\, \omega), \tag{15.13}$$

where $S_{s\ell m}$ are spin-weighted spheroidal harmonics and $R_{s\ell m}$ satisfies an ODE with regular singular points at the horizons and at infinity. From the perspective of geometry, this is a statement that Kerr spacetime admits enough structure to allow a clean spectral decomposition even without spherical symmetry.

Phenomenon statement 15.4: Frame dragging leaves fingerprints in radiation

Rotation is not just an extra parameter in a metric. In Kerr, frame dragging mixes time and azimuthal directions and produces an ergoregion where negative-energy states exist relative to infinity. In perturbation theory, this leads to superradiant amplification in certain frequency bands and modifies the quasinormal spectrum in a way that depends on the dimensionless spin a_*. Observationally, this is why ringdown frequencies and damping times encode spin so directly [5, 11].

15.5 Quasinormal modes: spectral geometry of horizons

15.5.1 Boundary conditions that create a discrete spectrum

A quasinormal mode (QNM) is not defined by regularity everywhere. It is defined by physically motivated boundary conditions: purely ingoing at the future event horizon and purely outgoing at future null infinity. These conditions represent the idea that the horizon absorbs and that no radiation comes in from infinity after the perturbation is set up. Under these constraints, the allowed frequencies become a discrete set of complex numbers $\omega_{\ell mn} = \omega_R + i\omega_I$ with $\omega_I < 0$ [9, 11]. The real part sets an oscillation frequency and the imaginary part sets an exponential decay time.

A useful phenomenological form for a single mode in the time domain is

$$h(t) \approx A\, e^{-t/\tau} \cos(2\pi f\, t + \phi), \tag{15.14}$$

where $f = \omega_R/(2\pi)$ and $\tau = -1/\omega_I$ are the mode frequency and damping time. The quality factor Q is defined by $Q = \pi f \tau$ and counts roughly how many oscillations occur before the amplitude decays significantly.

15.5.2 A practical fitting formula for the dominant Kerr mode

For the dominant gravitational mode $(\ell, m, n) = (2, 2, 0)$ of a Kerr black hole, the dimensionless complex frequency can be accurately approximated by fits of the form [10]

$$M\omega_R \approx f_1 + f_2(1 - a_*)^{f_3}, \tag{15.15}$$

$$Q \approx q_1 + q_2(1 - a_*)^{q_3}, \tag{15.16}$$

where M is the remnant mass, a_* is the dimensionless spin, and the coefficients $(f_1, f_2, f_3, q_1, q_2, q_3)$ are provided in [10]. These formulas are not a detector model. They are a compact representation of the Kerr perturbation spectrum and can be used to convert an inferred mass and spin into predicted ringdown frequency and damping time.

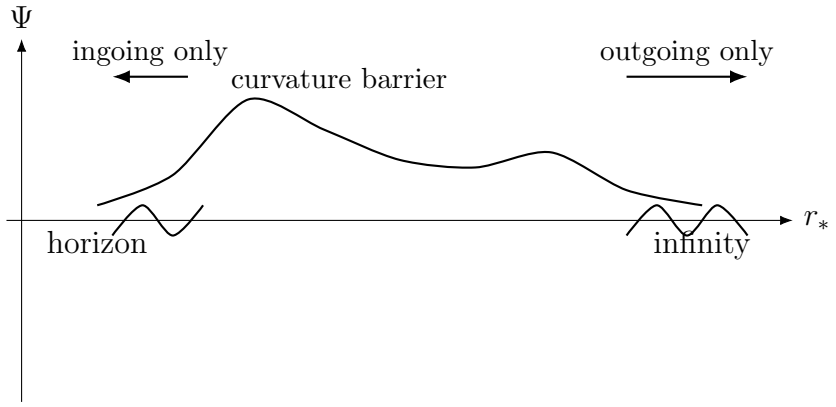


Figure 15.4: Conceptual diagram: quasinormal mode boundary conditions. The horizon condition is purely ingoing and the infinity condition is purely outgoing, producing a discrete set of complex eigenfrequencies.

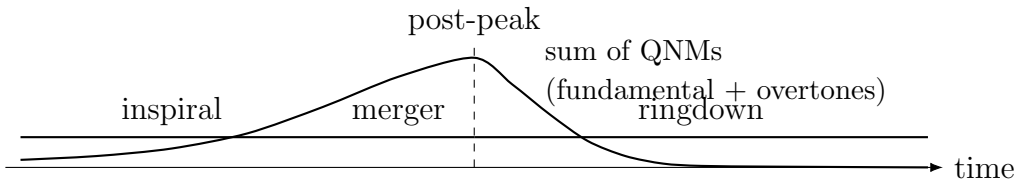


Figure 15.5: Conceptual diagram: inspiral-merger-ringdown as a geometric decomposition. The ringdown portion is modeled as a superposition of Kerr quasinormal modes whose frequencies and damping times are set by the remnant mass and spin.

Phenomenon statement 15.5: One event can already test the Kerr hypothesis

Ringdown-only analyses of GW150914 have reported evidence for the dominant QNM and at least one overtone. By measuring mass and spin using only post-peak data and checking consistency with the full inspiral-merger-ringdown inference, one can directly test whether the remnant behaves like a Kerr black hole, which is a concrete operational form of the no-hair idea [15, 16].

15.6 Data-driven illustrations from GWTC-1: from catalog values to ringdown scales

This section anchors the abstract perturbation discussion in open observational data. The dataset embedded (see `gwtc1_bbh.csv`) contains median parameter estimates for the ten confident binary black hole mergers in GWTC-1, extracted from the GWOSC event pages [1]. The parameters include component masses, chirp mass, effective inspiral spin, remnant mass and spin, radiated energy, peak luminosity, distance, redshift, sky

localization area, and a representative network matched filter SNR from the gstLAL pipeline.

Several additional columns are derived mathematically from these reported medians. In particular, the predicted dominant Kerr ringdown frequency and damping time are computed using the fitting formulas summarized in Section 15.4 and published in [10]. We also compute detector-frame versions by applying the redshift factors $f_{\mathrm{obs}} = f_{\mathrm{src}}/(1+z)$ and $\tau_{\mathrm{obs}} = (1+z)\tau_{\mathrm{src}}$.

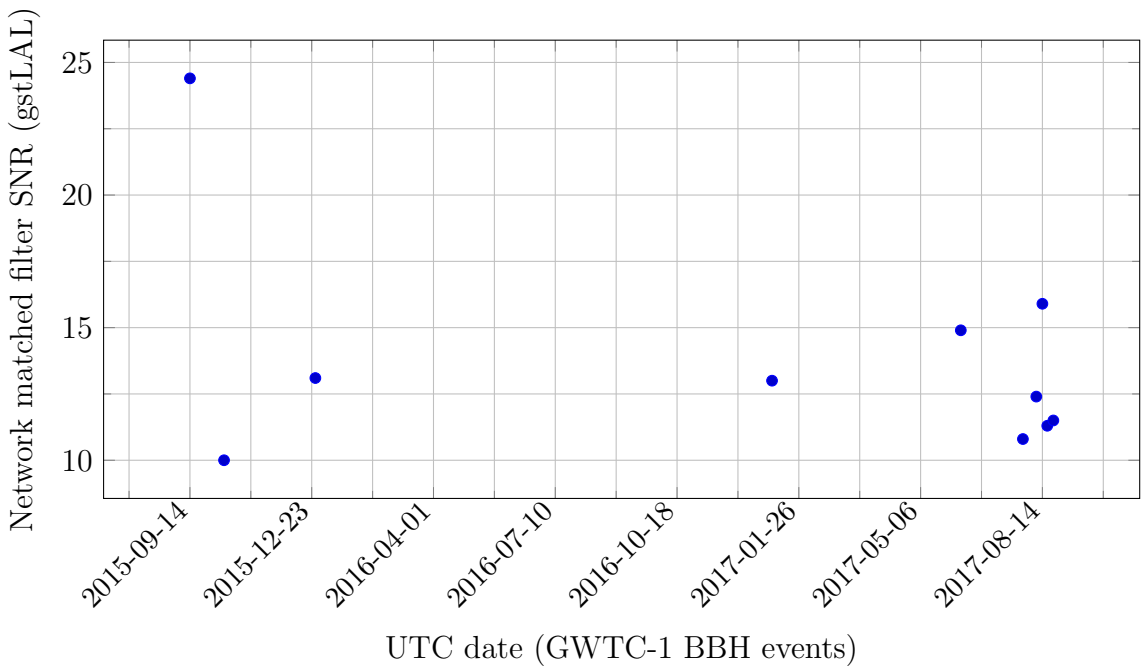


Figure 15.6: Data-based illustration 15.1: detection timeline and network matched filter SNR for the GWTC-1 binary black hole events used in this chapter. Higher SNR generally enables sharper constraints on late-time ringdown parameters. Data source: GWOSC GWTC-1 event pages [1].

15.6.1 Detection strength and observing time

15.6.2 Mass scale: chirp mass distribution

15.6.3 Spin information in the inspiral: effective spin versus mass ratio

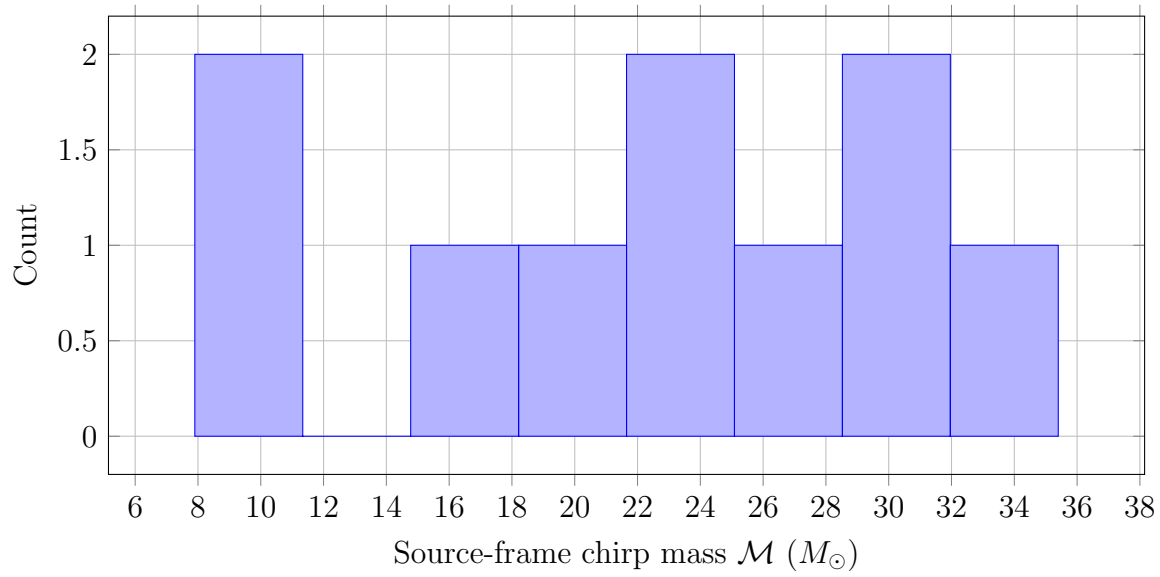


Figure 15.7: Data-based illustration 15.2: histogram of source-frame chirp masses for GWTC-1 binary black hole mergers. The chirp mass sets the dominant inspiral phase evolution and correlates with the final remnant mass that controls ringdown frequencies. Data source: GWOSC GWTC-1 event pages [1].

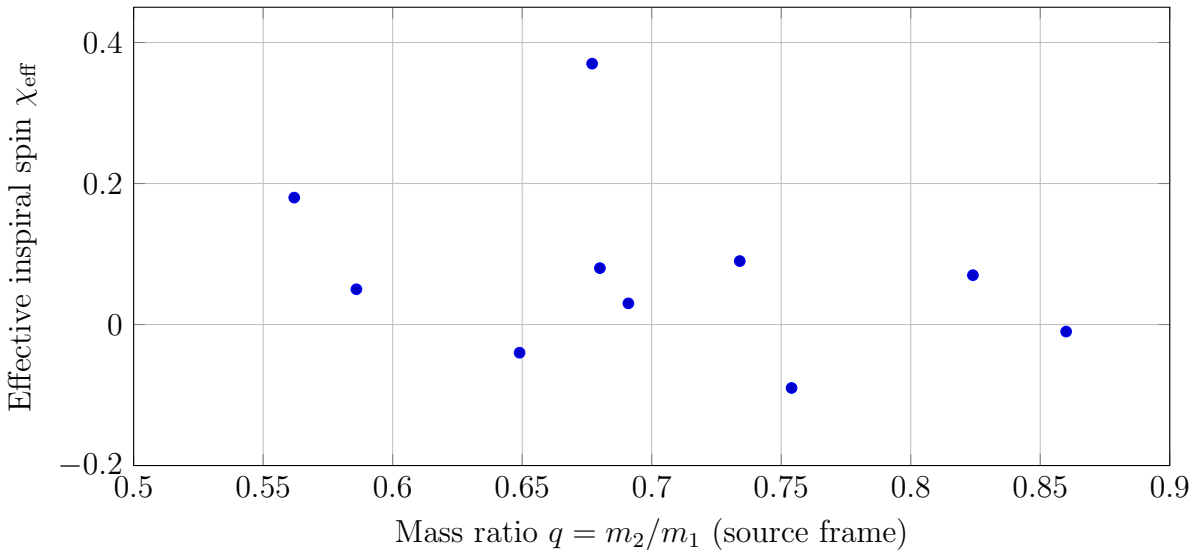


Figure 15.8: Data-based illustration 15.3: effective inspiral spin χ_{eff} versus mass ratio for GWTC-1 BBH events. The effective spin is a symmetry-adapted combination of component spins that survives strongly into the inspiral phase and influences the merger and remnant spin, which then shapes the ringdown spectrum. Data source: GWOSC [1].

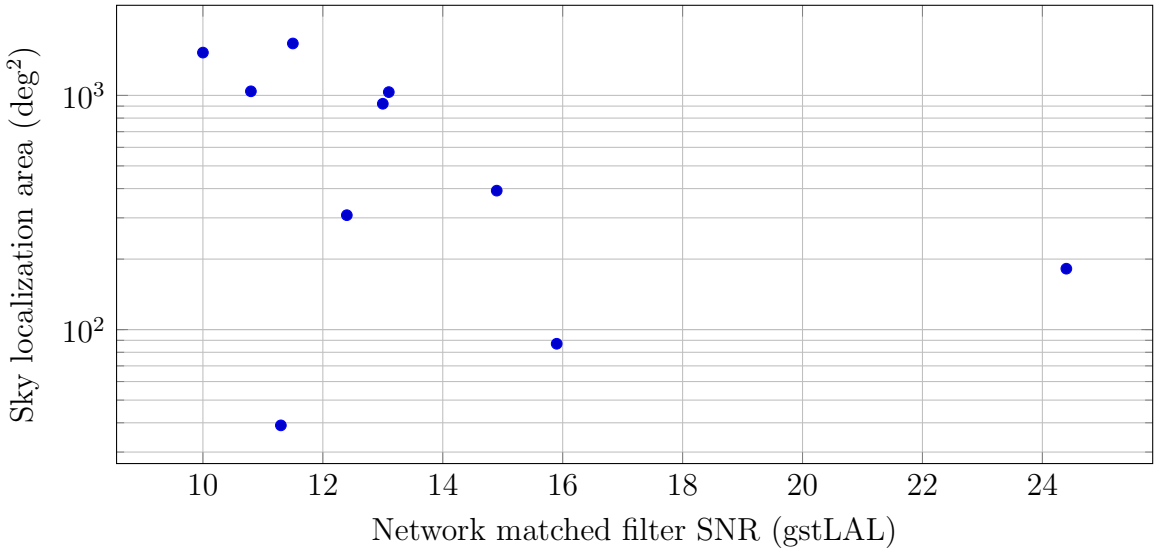


Figure 15.9: Data-based illustration 15.4: GWTC-1 sky localization area versus matched filter SNR. The relationship is not one-to-one because geometry matters: detector antenna patterns, polarization, and network baselines shape localization even at similar SNR. Data source: GWOSC [1].

15.6.4 Localization geometry: sky area versus SNR

15.6.5 Energy budget: radiated energy fraction

15.6.6 Peak luminosity scale: curvature radiation as a burst

15.6.7 Mass-energy closure check using catalog medians

15.6.8 Ringdown frequency scale from catalog remnant parameters

15.6.9 Cosmological redshift as a geometric rescaling of ringdown

15.6.10 Damping and quality: how spin controls decay

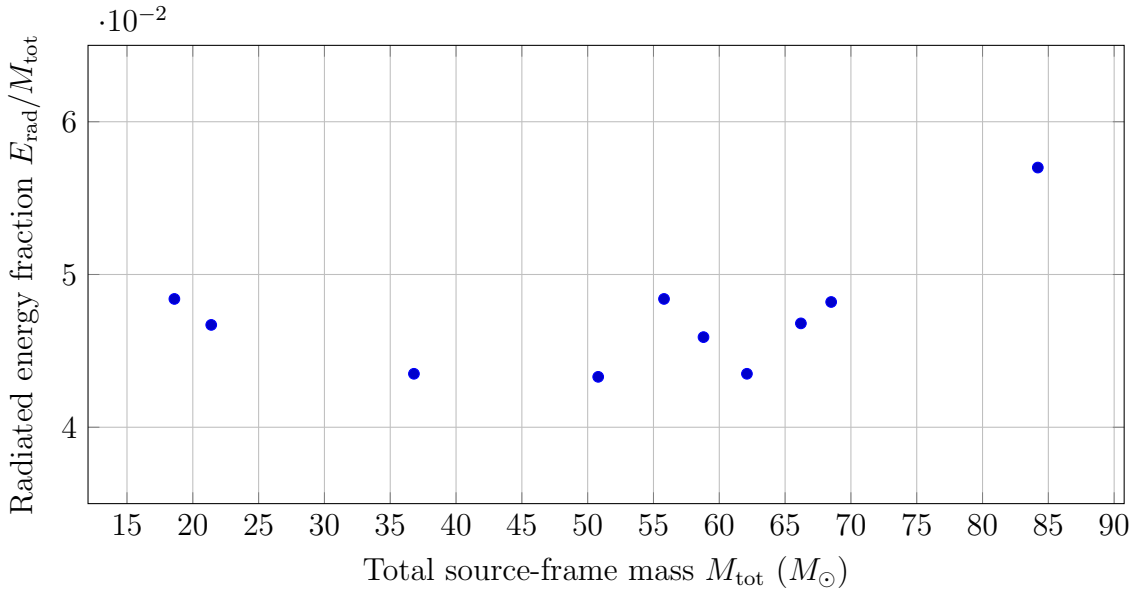

$\cdot 10^{-2}$
Radiated energy fraction $E_{\mathrm{rad}}/M_{\mathrm{tot}}$
6
5
4
15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90
Total source-frame mass M_{tot} $(M_\odot)$

Figure 15.10: Data-based illustration 15.5: fraction of total mass radiated in gravitational waves for GWTC-1 BBH mergers, using the reported radiated energy in units of $M_\odot c^2$. The near constancy around a few percent highlights a recurring geometric fact: mergers convert a modest fraction of mass into curvature radiation, and that conversion helps set the remnant parameters that perturbation theory then studies. Data source: GWOSC [1].

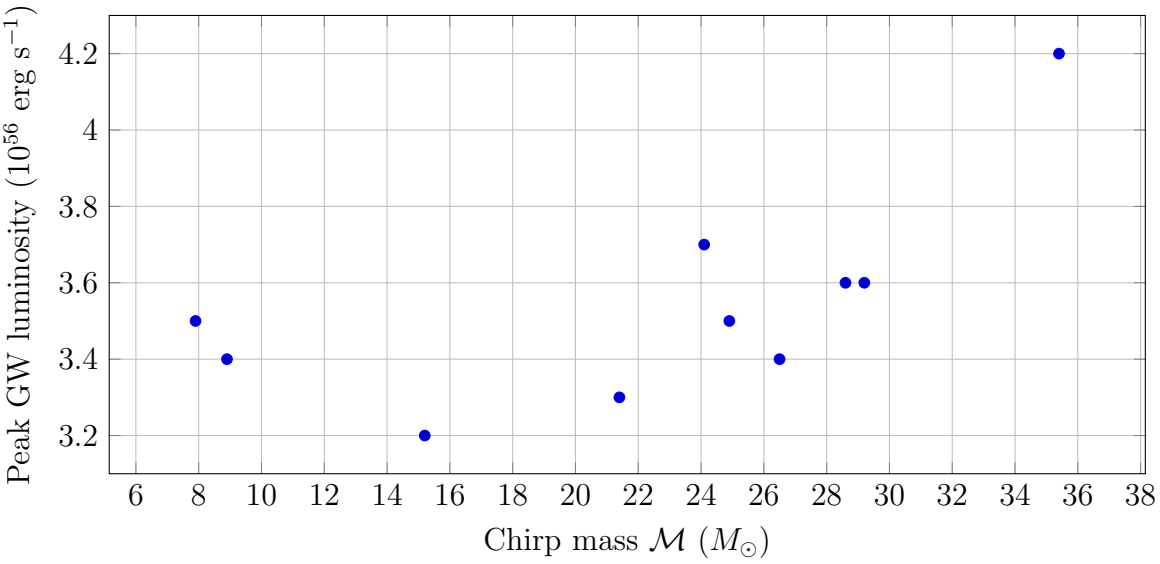


Figure 15.11: Data-based illustration 15.6: reported peak gravitational-wave luminosity versus chirp mass for GWTC-1 BBH events. Even though the luminosity peaks near merger and is not itself a ringdown quantity, it is part of the same perturbative story: the remnant is born in a violently distorted state and then relaxes via quasinormal emission. Data source: GWOSC [1].

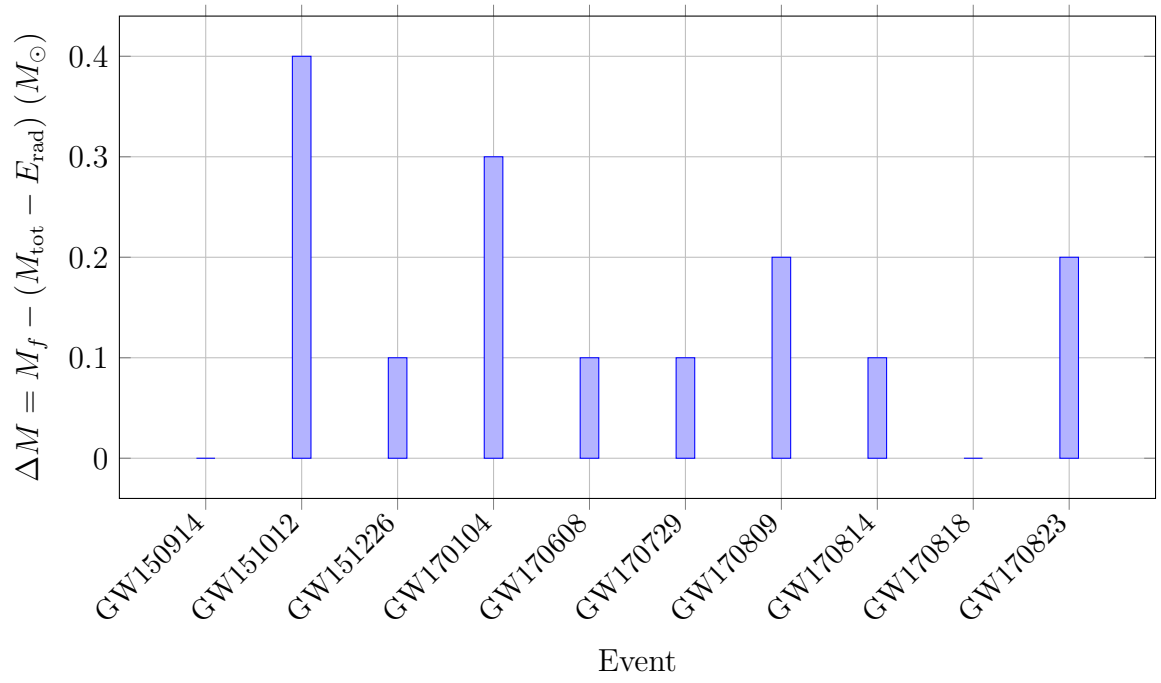


Figure 15.12: Data-based illustration 15.7: mass-energy closure check using catalog medians. Ideally, the remnant mass M_f equals the initial total mass minus radiated energy. Small nonzero residuals here reflect the fact that all quantities are posterior medians from parameter estimation rather than jointly constrained deterministic numbers. Data source: GWOSC [1].

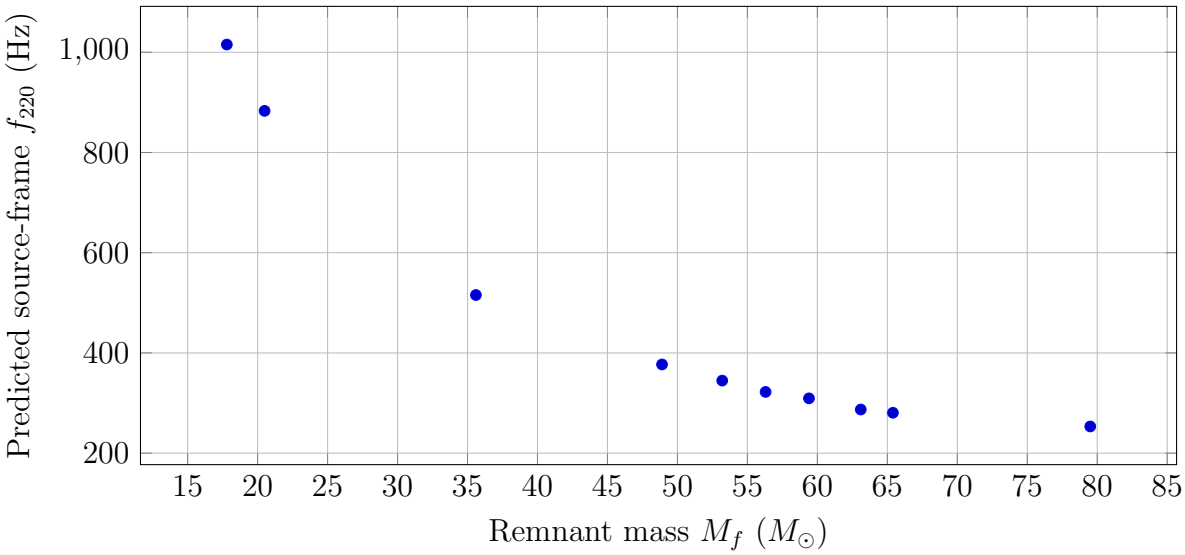


Figure 15.13: Data-based illustration 15.8: predicted source-frame dominant ringdown frequency f_{220} for GWTC-1 BBH remnants, derived from the reported remnant mass and spin using published Kerr QNM fitting formulas [10]. Data source for remnant parameters: GWOSC [1].

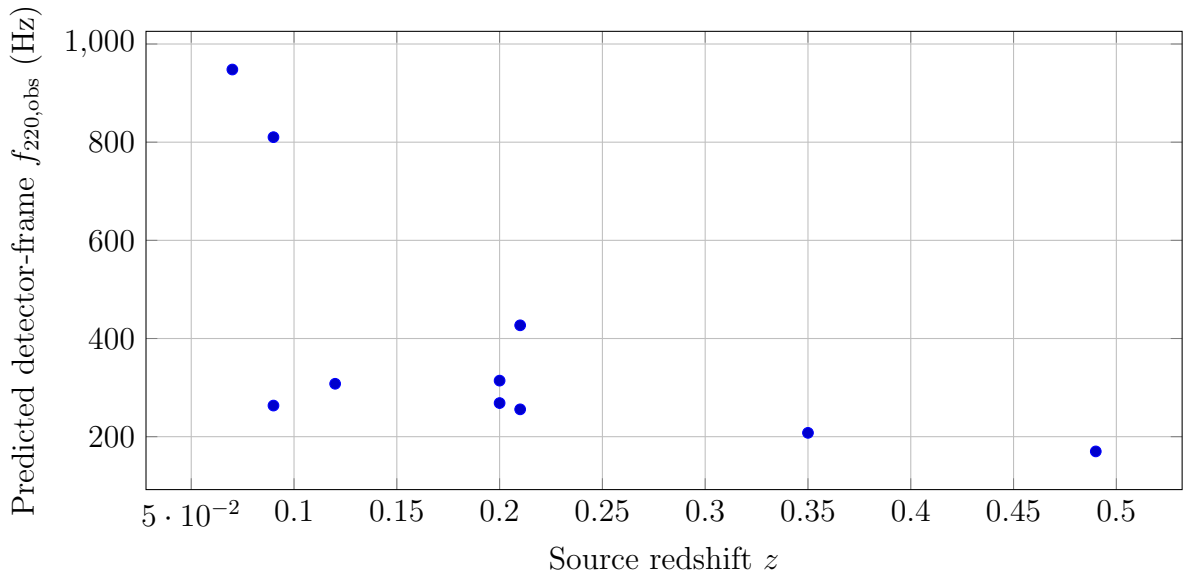


Figure 15.14: Data-based illustration 15.9: detector-frame dominant ringdown frequency after applying redshift, $f_{\mathrm{obs}} = f_{\mathrm{src}}/(1+z)$. In perturbation theory the intrinsic spectrum is geometric and local to the remnant, while observation adds the large-scale geometric effect of cosmological expansion. Data source: GWOSC [1].

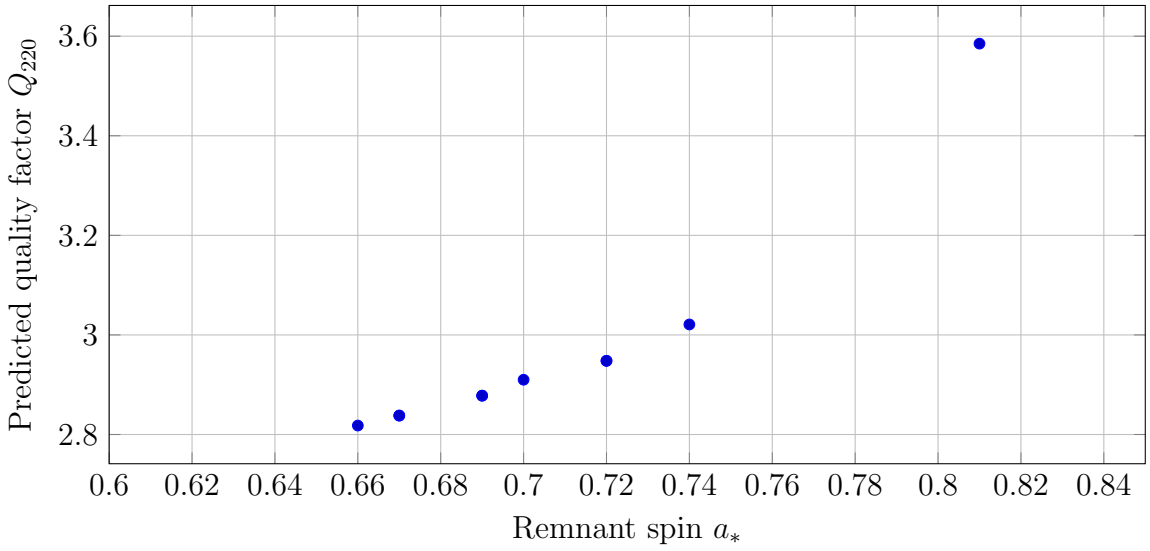


Figure 15.15: Data-based illustration 15.10: predicted quality factor $Q_{220} = \pi f \tau$ versus remnant spin. Higher spin typically leads to longer-lived oscillations (larger Q) for the dominant mode. Values are derived from catalog remnant parameters using fitting formulas [10]. Data source: GWOSC [1].

Phenomenon statement 15.6: Why these plots matter for observational inference

A perturbation equation is a promise that certain combinations of curvature behave like waves with precise boundary conditions. The GWTC-1 derived plots above show how that promise meets the detector band. For stellar-mass remnants, the dominant ringdown frequencies land in the tens to thousands of hertz range, exactly where ground-based interferometers are sensitive. For heavier remnants or higher redshift events, the same intrinsic spectrum is shifted downward. This is not a detail of instrumentation. It is the geometric statement that mass sets the time scale of spacetime response, and cosmological redshift rescales it again.

15.7 Stability: what "small ripples do not destroy the hole" really means

15.7.1 Mode stability, boundedness, decay, and nonlinear stability

The word stability is used casually, but in black hole perturbation theory it has a hierarchy of meanings. Mode stability asserts that there are no exponentially growing solutions of the separated form $e^{-i\omega t}$ with $\mathrm{Im}(\omega) > 0$ under physical boundary conditions. For Schwarzschild, the original Regge-Wheeler analysis and subsequent work established this structure [3, 8]. For Kerr, a major milestone is Whiting's proof of mode stability for the Teukolsky equation [12].

Boundedness and decay are stronger. They say that for generic initial data, the perturbation remains bounded and typically decays in time. For the scalar wave equation on Kerr backgrounds, definitive boundedness and decay results have been proven in the full subextremal regime [13, 14]. Although the full nonlinear stability of Kerr remains one of the deepest problems in mathematical relativity, the logic chain is clear: mode stability helps rule out spectral instabilities; energy estimates then control growth; decay estimates show relaxation; and only after this can one attempt a nonlinear bootstrap.

15.7.2 Interpretation through horizons and causal structure

The horizon matters for stability in two intertwined ways. First, it sets the boundary condition: perturbations can fall in, but classical horizons do not let information escape back out. Second, it reshapes causal propagation. Because the horizon is a null hypersurface, the natural energy currents used in estimates must be carefully adapted to the causal geometry. This is why stability is not just a PDE problem but a geometric one.

From an observational viewpoint, stability is the reason ringdown exists at all as a clean

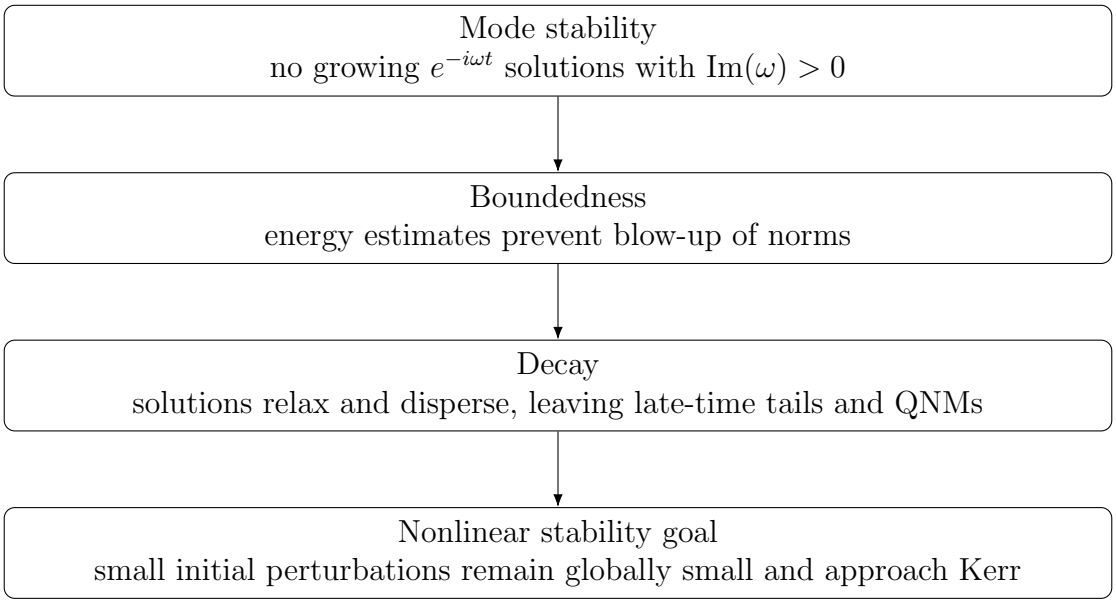

Mode stability
no growing $e^{-i\omega t}$ solutions with $\mathrm{Im}(\omega) > 0$
Boundedness
energy estimates prevent blow-up of norms
Decay
solutions relax and disperse, leaving late-time tails and QNMs
Nonlinear stability goal
small initial perturbations remain globally small and approach Kerr

Figure 15.16: Conceptual diagram: the stability ladder in black hole perturbation theory. Each rung is a stronger statement about dynamics, and the full nonlinear question is far more demanding than the absence of unstable separated modes.

signal component. If small perturbations grew, the late-time waveform would not settle into a predictable damped oscillation. The empirical fact that observed ringdowns are consistent with decaying QNMs is, therefore, evidence supporting the geometric stability picture predicted by perturbation theory [15, 17].

Phenomenon statement 15.7: Why overtone modeling changes the geometric narrative

In a naive story, ringdown begins only after the spacetime becomes sufficiently close to Kerr. Overtone analyses complicate this story in a useful way: by including higher overtones, one can model the post-peak signal earlier, suggesting that the geometry behaves like a superposition of Kerr QNMs sooner than expected [16]. The meaning is geometric: the QNM basis can approximate the near-horizon curvature response even when the perturbation amplitude is not yet tiny, although the interpretation becomes more subtle and depends on how one defines "close to Kerr".

15.8 Conclusion

Perturbations of black hole geometry are not a technical appendix to exact solutions. They are the mathematical engine that converts curvature into observable predictions. The linearized Einstein equations define a wave operator on tensor fields with built-in gauge redundancy. In Schwarzschild, symmetry compresses the problem into the Regge-Wheeler and Zerilli master equations, revealing curvature potentials that scatter waves and

generate universal ringdown patterns. In Kerr, the correct language is curvature-based: Newman-Penrose scalars satisfy the Teukolsky equation, enabling separability and direct access to radiative degrees of freedom.

Quasinormal modes unify these perspectives by turning horizon boundary conditions into a spectral problem. Their discrete complex frequencies are geometric invariants of the remnant that can be inferred from gravitational-wave data. The data-driven figures built from GWTC-1 show how catalog remnant parameters translate, through published QNM fits, into ringdown frequencies and quality factors that sit in the sensitive band of ground-based detectors. Finally, the notion of stability emerges as a hierarchy, from mode stability through decay to the deeper nonlinear questions. The lesson is that "small ripples do not destroy the hole" is a precise statement about the way horizons, curvature, and causal structure constrain the dynamics of spacetime itself.

Bibliography

[1] Gravitational Wave Open Science Center (GWOSC), "GWTC-1 event portal and open data release," DOI: 10.7935/82H3-HH23. [Online]. Available: https://gwosc.org/eventapi/html/GWTC-1-confident/

[2] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs," *Phys. Rev. X*, vol. 9, no. 3, p. 031040, 2019, doi: 10.1103/PhysRevX.9.031040. [Online]. Available: https://arxiv.org/abs/1811.12907

[3] T. Regge and J. A. Wheeler, "Stability of a Schwarzschild singularity," *Phys. Rev.*, vol. 108, pp. 1063-1069, 1957, doi: 10.1103/PhysRev.108.1063. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.108.1063

[4] F. J. Zerilli, "Gravitational field of a particle falling in a Schwarzschild geometry analyzed in tensor harmonics," *Phys. Rev. D*, vol. 2, pp. 2141-2160, 1970, doi: 10.1103/PhysRevD.2.2141. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.2.2141

[5] S. A. Teukolsky, "Perturbations of a rotating black hole. I. Fundamental equations for gravitational, electromagnetic, and neutrino-field perturbations," *Astrophys. J.*, vol. 185, pp. 635-647, 1973, doi: 10.1086/152444. [Online]. Available: https://www.osti.gov/biblio/4318110

[6] E. Newman and R. Penrose, "An approach to gravitational radiation by a method of spin coefficients," *J. Math. Phys.*, vol. 3, no. 3, pp. 566-578, 1962, doi: 10.1063/1.1724257. [Online]. Available: https://pubs.aip.org/aip/jmp/article/3/3/566/387969/An-Approach-to-Gravitational-Radiation-by-a-Method

[7] K. Martel and E. Poisson, "Gravitational perturbations of the Schwarzschild spacetime: A practical covariant and gauge-invariant formalism," 2005. [Online]. Available: https://arxiv.org/abs/gr-qc/0502028

[8] C. V. Vishveshwara, "Stability of the Schwarzschild metric," *Phys. Rev. D*, vol. 1, pp. 2870-2879, 1970, doi: 10.1103/PhysRevD.1.2870. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.1.2870

[9] K. D. Kokkotas and B. G. Schmidt, "Quasi-normal modes of stars and black holes," *Living Rev. Relativity*, vol. 2, no. 2, 1999, doi: 10.12942/lrr-1999-2. [Online]. Available: https://arxiv.org/abs/gr-qc/9909058

[10] E. Berti, V. Cardoso, and C. M. Will, "Gravitational-wave spectroscopy of massive black holes with the space interferometer LISA," *Phys. Rev. D*, vol. 73, p. 064030, 2006, doi: 10.1103/PhysRevD.73.064030. [Online]. Available: https://arxiv.org/abs/gr-qc/0512160

[11] E. Berti, V. Cardoso, and A. O. Starinets, "Quasinormal modes of black holes and black branes," *Class. Quantum Grav.*, vol. 26, p. 163001, 2009, doi: 10.1088/0264-9381/26/16/163001. [Online]. Available: https://arxiv.org/abs/0905.2975

[12] B. F. Whiting, "Mode stability of the Kerr black hole," *J. Math. Phys.*, vol. 30, no. 6, pp. 1301-1305, 1989, doi: 10.1063/1.528308. [Online]. Available: https://pubs.aip.org/aip/jmp/article/30/6/1301/228680/Mode-stability-of-the-Kerr-black-hole

[13] M. Dafermos, I. Rodnianski, and Y. Shlapentokh-Rothman, "Decay for solutions of the wave equation on Kerr exterior spacetimes III: The full subextremal case $|a| < M$," 2014. [Online]. Available: https://arxiv.org/abs/1402.7034

[14] M. Dafermos, I. Rodnianski, and Y. Shlapentokh-Rothman, "Decay for solutions of the wave equation on Kerr exterior spacetimes III: The full subextremal case $|a| < M$," *Ann. of Math.*, vol. 183, no. 3, pp. 787-913, 2016. [Online]. Available: https://annals.math.princeton.edu/2016/183-3/p02

[15] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, "Testing the no-hair theorem with GW150914," *Phys. Rev. Lett.*, vol. 123, p. 111102, 2019, doi: 10.1103/PhysRevLett.123.111102. [Online]. Available: https://arxiv.org/abs/1905.00869

[16] M. Giesler, M. Isi, M. A. Scheel, and S. A. Teukolsky, "Black hole ringdown: The importance of overtones," *Phys. Rev. X*, vol. 9, p. 041060, 2019, doi: 10.1103/PhysRevX.9.041060. [Online]. Available: https://arxiv.org/abs/1903.08284

[17] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1," *Phys. Rev. D*, vol. 100, p. 104036, 2019, doi: 10.1103/PhysRevD.100.104036. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.100.104036

[18] A. Pound and B. Wardell, "Black hole perturbation theory and gravitational self-force," 2021. [Online]. Available: https://arxiv.org/abs/2101.04592

[19] E. W. Leaver, "An analytic representation for the quasi-normal modes of Kerr black holes," *Proc. R. Soc. A*, vol. 402, no. 1823, pp. 285-298, 1985, doi: 10.1098/rspa.1985.0119. [Online]. Available: https:

//royalsocietypublishing.org/rspa/article/402/1823/285/15806/
An-analytic-representation-for-the-quasi-normal

Black Holes Geometry: Curvature and Horizons

Chapter 16: Optical Geometry, Lensing, and Finsler Randers Tricks

Contents

IV Geometric Methods for Dynamics and Perturbations **5**

16 Optical Geometry, Lensing, and Finsler Randers Tricks **6**

16.1 Introduction . . . 6
16.2 Null geodesics as optics in curved spacetime . . . 7
16.2.1 From wavefronts to rays . . . 7
16.2.2 Conceptual diagram: lensing geometry and angles . . . 8
16.3 Stationary spacetimes and Fermat's principle . . . 8
16.3.1 Stationary decomposition and arrival time . . . 8
16.3.2 Conceptual diagram: Fermat potential and stationary points . . . 9
16.3.3 Conceptual diagram: optical projection of null geodesics . . . 9
16.4 From Fermat principle to the thin lens equation . . . 9
16.5 Optical metric in static spacetimes . . . 10
16.5.1 Static case and Riemannian optical geometry . . . 10
16.5.2 Conceptual diagram: Gaussian curvature and bending as a global integral 11
16.5.3 Gauss Bonnet deflection as a curvature integral . . . 11
16.6 Randers and Finsler viewpoint for rotating spacetimes . . . 12
16.6.1 Randers form and the wind one form . . . 12
16.6.2 Explicit Randers data from metric components . . . 12
16.6.3 Conceptual diagram: Randers navigation picture . . . 12
16.7 Photon spheres, photon regions, and shadow boundaries . . . 13
16.7.1 Shadow as a set of unstable null geodesics . . . 13
16.7.2 Conceptual diagram: photon sphere and shadow edge . . . 13
16.8 Data driven optical geometry in strong lens catalogs: SLACS case study . . . 14
16.8.1 Dataset provenance and why it matters . . . 14
16.8.2 Einstein radius distribution as an optical observable . . . 14
16.8.3 Lens and source redshifts: the distance lever arm . . . 14
16.8.4 Data test of the singular isothermal sphere scaling . . . 15
16.8.5 How far from SIS: distribution of the ratio $\sigma/\sigma_{\rm SIS}$ . . . 16
16.8.6 From angular optics to enclosed mass: Einstein mass distribution . . . 16
16.8.7 Einstein mass versus stellar mass: where optics meets baryons . . . 17
16.8.8 Where lensing lives inside the galaxy: $R_E/R_{\rm eff}$ . . . 17
16.9 Historical and experimental anchors for the optical viewpoint . . . 18
16.9.1 1919 eclipse report: early data on gravitational deflection . . . 18
16.9.2 Gravity Probe B: frame dragging as the Randers drift sector . . . 19
16.9.3 Event Horizon Telescope: ring diameters and the shadow boundary . . . 19
16.10Phenomenon statements and real world applications . . . 21

16.11Conclusion 22

List of Figures

16.1 Conceptual lensing geometry showing observer O, lens L, true source position (unlensed) S', and lensed source position S with observed angle θ and true angle β. This diagram fixes notation used in the rest of the chapter. 8
16.2 Conceptual Fermat potential surface. Lensed images form at stationary points of the arrival time functional, and relative time delays between images are differences of this functional evaluated at those stationary points. This picture operationalizes Fermat's principle in lens modeling [3, 14]. 9
16.3 Conceptual optical geometry: a null geodesic in spacetime projects to a curve on a spatial slice. Optical geometry constructs a metric or Finsler structure on the slice so that this projection is geodesic with respect to that structure. 9
16.4 Conceptual Gauss Bonnet domain for lensing. In optical geometry approaches the deflection angle can be expressed using an integral of Gaussian curvature K over a domain $\mathcal{D}$ bounded by the ray and auxiliary curves, plus boundary terms. 11
16.5 Conceptual Randers picture: a base Riemannian norm α sets an isotropic local cost, while a one form term β acts like a drift. In stationary spacetimes β is built from g_{ti} and encodes frame dragging in the optical problem [5]. 13
16.6 Conceptual relation between unstable photon orbits and the shadow boundary. The shadow edge is defined by those null geodesics that asymptotically approach unstable photon orbits before reaching the observer. Observationally, the Event Horizon Telescope constrains this boundary through ring diameter measurements [10, 11]. 13
16.7 Histogram of Einstein radii θ_E from the debiased SLACS strong lens sample [1]. θ_E is a directly observed optical scale that sets the size of the critical curve in the image plane. 15
16.8 Lens redshift z_d versus source redshift z_s for the SLACS debiased sample [1]. The diagonal line marks $z_s = z_d$; physical systems lie above it. The spread controls the distance ratio lever arm that enters optical lensing formulas. 15
16.9 Observed stellar velocity dispersion σ compared with $\sigma_{\rm SIS}$ inferred by inverting the SIS Einstein radius relation using (θ_E, z_d, z_s) for each SLACS system [1]. Distances are computed in a flat ΛCDM fiducial cosmology following [2]. 16
16.10Histogram of the ratio $\sigma/\sigma_{\rm SIS}$ across the SLACS sample [1]. The vertical reference at 1 indicates exact SIS scaling consistency. 16
16.11Histogram of $\log_{10} M_E$ derived from SLACS measured (θ_E, z_d, z_s) using Eq. (16.17) with angular diameter distances computed following [2]. Input lens data are from [1]. 17

400

16.12Derived Einstein mass M_E versus catalog stellar mass $M_\star$ for SLACS systems [1]. Distances follow [2]. The diagonal is a visual reference. 18
16.13Ratio of physical Einstein radius to effective radius, R_E/R_{eff}, versus stellar mass for SLACS systems [1]. This ratio describes how deeply the optical critical curve probes the lens galaxy relative to its light distribution. 18
16.14Limb deflection estimates stated in the 1919 eclipse expedition report for the astrographic plates [7], compared with Einstein and Newton benchmark values. The spread illustrates how an optical invariant can be sensitive to reduction assumptions when data quality is marginal. 19
16.15Gravity Probe B predicted and measured drift rates for the geodetic and frame dragging effects [8, 9]. The frame dragging sector corresponds to the stationary off diagonal structure that becomes the Randers drift one form in optical geometry. 20
16.16Ring diameter constraints reported in Event Horizon Telescope related analyses for M87 and Sgr A [10, 11]. These angular scales constrain the shadow boundary determined by unstable photon orbits. 20

Part IV

Geometric Methods for Dynamics and Perturbations

Chapter 16

Optical Geometry, Lensing, and Finsler Randers Tricks

Abstract

Optical geometry is the program of rewriting light propagation in curved spacetime as an effective geometry problem on a spatial manifold. When this rewriting is done carefully, null geodesics become geodesics of an optical metric in static spacetimes and become geodesics of a Randers type Finsler structure in stationary spacetimes. This chapter develops the geometric machinery behind that statement and then anchors it to real observational and experimental datasets: strong lens systems from the debiased SLACS catalog, historical solar eclipse deflection estimates from the 1919 expedition report, the Gravity Probe B measurement of frame dragging, and Event Horizon Telescope constraints on ring diameters that define the shadow boundary. The main message is that the optical viewpoint is not only a change of variables but also a bridge between curvature, causality, and measurement. It packages lensing, time delays, and shadow edges into geometric invariants that can be compared directly to data.

Keywords

Optical metric; Fermat principle; gravitational lensing; time delay; Gauss Bonnet theorem; Finsler geometry; Randers metric; frame dragging; photon region; black hole shadow; strong lensing data.

16.1 Introduction

Light is the natural probe of geometry because its worldlines are null. In general relativity, that means photons trace null geodesics, and null geodesics are determined entirely by the metric up to conformal rescaling. This simple sentence hides a powerful operational idea: if we can observe enough light paths, we can reconstruct geometric information about spacetime without ever touching the spacetime directly. Gravitational lensing is the clearest instantiation of this idea. A lens system provides multiple images, image separations, time delays, magnification patterns, and sometimes polarization rotation, all produced by the same underlying causal structure. The modern black hole shadow is a similarly geometric observable: it is a boundary on the observer's sky that separates captured null geodesics from escaping null geodesics. That boundary is

determined by unstable photon orbits, and therefore by the curvature near the horizon, yet it is reported to us as a two dimensional shape in data.

The technical challenge is that null geodesics live in four dimensions, while most lensing intuition and almost all data analysis occur on two or three dimensional spaces: the observer's screen, the lens plane, and projected sky coordinates. Optical geometry is the systematic answer. It constructs an effective geometry in which the spatial projection of light rays is itself geodesic motion, so that the tools of Riemannian geometry can be applied directly to lensing. In static spacetimes this optical geometry is Riemannian; in stationary spacetimes it is generically Finslerian, and in a particularly tractable way it becomes a Randers metric. The Randers viewpoint makes frame dragging behave like a drift one form, which is why it is often described with the language of winds and navigation.

This chapter is written with two goals. The first is conceptual: to make the optical and Randers constructions feel inevitable rather than mysterious, and to emphasize which assumptions are doing the work. The second is empirical: to ensure that the central quantities are not left floating as symbols. For that reason, the illustrations are data based whenever possible. The SLACS strong lens sample supplies measured Einstein radii, redshifts, and velocity dispersions, which allow us to test the simplest optical model, the singular isothermal sphere, directly against a real catalog. The 1919 eclipse report supplies historical deflection estimates, which help interpret why lensing is subtle when instruments and atmospheres intervene. Gravity Probe B supplies a clean measurement of the gravitomagnetic sector of the metric, which is the same sector that appears as the Randers one form. Finally, the Event Horizon Telescope supplies ring diameters for M87 and Sgr A, which constrain the geometry of the photon region that defines the shadow boundary.

Reading map. Sections 2 to 4 build the core optical geometry construction in a coordinate friendly but coordinate aware way. Section 5 explains why rotating spacetimes push us into Randers and Finsler geometry. Section 6 connects this to photon spheres and shadows. Sections 7 and 8 are data driven and include the required analysis figures. Section 9 lists application oriented phenomenon statements, emphasizing what the optical viewpoint lets you compute and measure.

16.2 Null geodesics as optics in curved spacetime

16.2.1 From wavefronts to rays

In geometric optics, light is described by an eikonal function $S(x)$ whose level sets $S = \text{const}$ represent wavefronts. The wave covector $k_a = \nabla_a S$ satisfies the eikonal equation

$$g^{ab} k_a k_b = 0, \tag{16.1}$$

which says that k^a is null. The corresponding integral curves of k^a are null geodesics up to affine reparametrization. In this view, curvature acts on light not as a force but as a constraint on which gradients can remain null. A useful operational point is that the overall scaling of k_a does not change the ray direction, which is the geometric reason lensing is conformally invariant.

A ray description becomes most transparent when written as a variational problem. For a

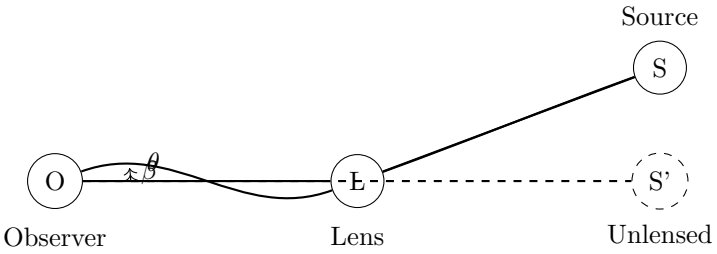


Figure 16.1: Conceptual lensing geometry showing observer O, lens L, true source position (unlensed) S', and lensed source position S with observed angle θ and true angle β. This diagram fixes notation used in the rest of the chapter.

null curve γ, parametrized by λ, one can consider the action

$$\mathcal{S}[\gamma] = \int L\, d\lambda, \qquad L = \frac{1}{2} g_{ab} \dot{x}^a \dot{x}^b, \tag{16.2}$$

with the null condition $L = 0$ imposed. The Euler Lagrange equations produce the geodesic equation $\dot{x}^b \nabla_b \dot{x}^a = 0$. This is a four dimensional statement. The optical idea is to rewrite it so that the spatial projection of γ satisfies a lower dimensional geodesic equation in a geometry that can be studied with standard tools.

16.2.2 Conceptual diagram: lensing geometry and angles

Figure 16.1 summarizes the basic observational angles in gravitational lensing. This is a conceptual diagram that fixes notation rather than supplying data.

16.3 Stationary spacetimes and Fermat's principle

16.3.1 Stationary decomposition and arrival time

A spacetime is stationary if it admits a timelike Killing vector field ξ^a that represents time translation symmetry. In adapted coordinates (t, x^i) where $\xi^a = (\partial_t)^a$, the metric can be written as

$$ds^2 = g_{tt} dt^2 + 2 g_{ti} dt\, dx^i + g_{ij} dx^i dx^j, \tag{16.3}$$

with all components independent of t. For null curves $ds^2 = 0$ this equation can be solved for dt in terms of the spatial increment dx^i. The result is a travel time functional

$$T[\gamma] = \int_\gamma dt = \int_\gamma \left(-\frac{g_{ti}}{g_{tt}} dx^i + \sqrt{\frac{g_{ij}}{-g_{tt}} dx^i dx^j + \frac{g_{ti} g_{tj}}{g_{tt}^2} dx^i dx^j} \right), \tag{16.4}$$

where the square root is chosen so that $dt > 0$ for future directed rays. Fermat's principle in stationary spacetimes states that the physical light path between fixed endpoints extremizes this arrival time functional. A detailed spacetime oriented discussion of this principle, including when it holds globally and how caustics enter, is reviewed by Perlick [3].

Equation (16.4) already contains the seed of Finsler geometry. The integrand is homogeneous of degree one in the spatial tangent vector, not quadratic, which is why the resulting geometry is not strictly Riemannian. When $g_{ti} = 0$ (static spacetimes) the first term vanishes and the square root becomes the length element of a genuine Riemannian optical metric. When $g_{ti} \neq 0$

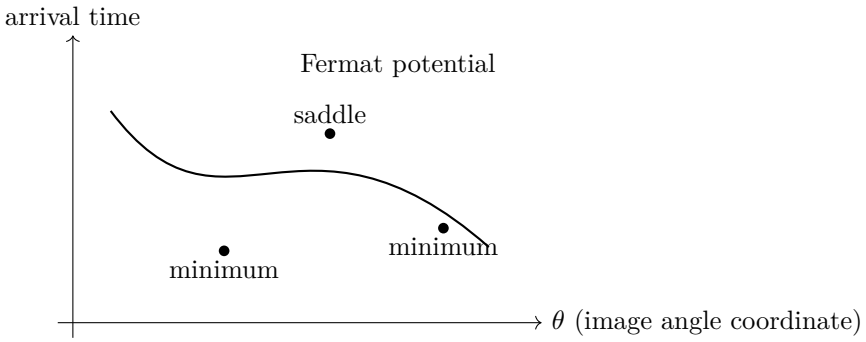


Figure 16.2: Conceptual Fermat potential surface. Lensed images form at stationary points of the arrival time functional, and relative time delays between images are differences of this functional evaluated at those stationary points. This picture operationalizes Fermat's principle in lens modeling [3, 14].

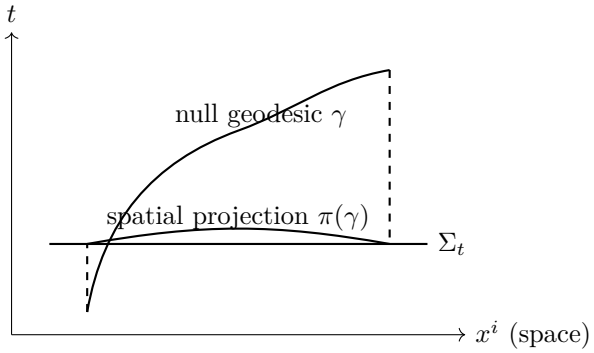


Figure 16.3: Conceptual optical geometry: a null geodesic in spacetime projects to a curve on a spatial slice. Optical geometry constructs a metric or Finsler structure on the slice so that this projection is geodesic with respect to that structure.

(rotating or more generally stationary non static spacetimes) the first term acts like a drift, and the square root plus drift is precisely of Randers form.

16.3.2 Conceptual diagram: Fermat potential and stationary points

In lensing theory, Fermat's principle is often visualized through an effective time delay surface, also called the Fermat potential. Images appear at stationary points of this surface: minima, saddles, and maxima. This conceptual picture is the reason that lensing can produce multiple images, and it is also the reason that image time delays are naturally interpreted as differences of a geometric functional rather than as arbitrary astrophysical offsets.

16.3.3 Conceptual diagram: optical projection of null geodesics

Figure 16.3 sketches the idea of projecting null geodesics to a spatial manifold where they become geodesics of an optical structure.

16.4 From Fermat principle to the thin lens equation

The Fermat functional in Eq. (16.4) is a fully relativistic statement, but most lensing data analysis is performed in a thin lens approximation. Optical geometry provides a clean bridge

between the two. In the weak field regime, write the metric in Newtonian gauge with potential Φ and ignore higher order terms. One obtains an effective refractive index

$$n(\mathbf{x}) \approx 1 - \frac{2\Phi(\mathbf{x})}{c^2}, \tag{16.5}$$

so that the travel time is approximately the Euclidean length weighted by $n(\mathbf{x})$. The deflection angle is then the integrated transverse gradient of Φ along the unperturbed path. This is the optical derivation of gravitational bending used in lensing theory texts [14, 3].

In the thin lens approximation, all bending is treated as occurring at a lens plane. The mapping between the true source angle β and the observed image angle θ is

$$\beta = \theta - \frac{D_{ds}}{D_s}\,\hat{\alpha}(\theta), \tag{16.6}$$

where $\hat{\alpha}(\theta)$ is the reduced deflection angle. In many models $\hat{\alpha}$ is expressed as the gradient of a scalar lensing potential $\psi(\theta)$, $\hat{\alpha} = \nabla_\theta \psi$. The arrival time for a light ray can be written as a Fermat potential

$$t(\theta, \beta) = \frac{1 + z_d}{c}\frac{D_d D_s}{D_{ds}}\left[\frac{1}{2}|\theta - \beta|^2 - \psi(\theta)\right], \tag{16.7}$$

up to an additive constant. Taking a gradient with respect to θ and setting it to zero yields the lens equation (16.6). This shows explicitly how lensing observables arise from the stationarity of an optical functional, which is the thin lens manifestation of Fermat's principle.

This viewpoint also clarifies why time delays are not optional complications. They are differences of the same geometric functional evaluated at different stationary points. If a variable source produces multiple images, the light curves should repeat with a delay Δt that probes the distance scale factor $D_d D_s / D_{ds}$ and the potential term ψ. That is why time delay cosmography is feasible in principle: the optical functional links geometry to observables in a way that is sensitive to cosmological distance scaling.

16.5 Optical metric in static spacetimes

16.5.1 Static case and Riemannian optical geometry

If the spacetime is static, the cross terms vanish: $g_{ti} = 0$. Write

$$ds^2 = -V^2(x)\,dt^2 + h_{ij}(x)\,dx^i dx^j. \tag{16.8}$$

For a null curve, $ds^2 = 0$ implies

$$dt = \frac{1}{V(x)}\sqrt{h_{ij}dx^i dx^j}. \tag{16.9}$$

Therefore the arrival time is the Riemannian length of the spatial curve with respect to the optical metric

$$\bar{h}_{ij} = V^{-2} h_{ij}. \tag{16.10}$$

Light rays projected to the spatial manifold are geodesics of $\bar{h}_{ij}$. This construction is deceptively strong: it turns the lensing deflection problem into a question about curvature of $\bar{h}_{ij}$ and about global topology of the domain through which the ray passes. That bridge is exploited explicitly

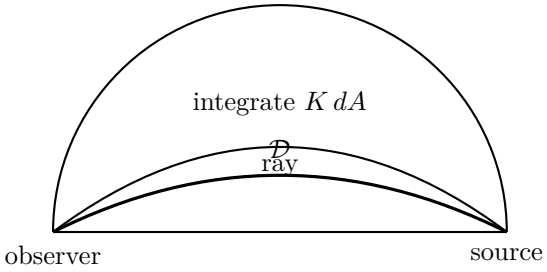


Figure 16.4: Conceptual Gauss Bonnet domain for lensing. In optical geometry approaches the deflection angle can be expressed using an integral of Gaussian curvature K over a domain $\mathcal{D}$ bounded by the ray and auxiliary curves, plus boundary terms.

in the Gauss Bonnet approach to lensing developed by Gibbons and Werner [4].

16.5.2 Conceptual diagram: Gaussian curvature and bending as a global integral

Figure 16.4 illustrates the geometric idea behind Gauss Bonnet lensing methods: the deflection angle can be related to an integral of Gaussian curvature over a region bounded by the light ray and suitable auxiliary curves.

16.5.3 Gauss Bonnet deflection as a curvature integral

The reason optical geometry is powerful is that it permits global statements. In a two dimensional optical manifold (for example, the equatorial plane of a static spherically symmetric spacetime after suitable reduction), the Gauss Bonnet theorem relates the integral of Gaussian curvature over a domain to boundary curvature and corner angles. Gibbons and Werner showed that, for an asymptotically flat optical geometry and a suitable domain bounded by the light ray and auxiliary curves, the total deflection angle $\hat{\alpha}$ can be expressed schematically as

$$\hat{\alpha} = -\iint_{\mathcal{D}} K\,dA + \int_{\partial\mathcal{D}} \kappa_g\,ds + \sum_i \theta_i - \pi, \tag{16.11}$$

where K is the Gaussian curvature of the optical metric, κ_g is the geodesic curvature of boundary segments that are not geodesics, and the θ_i are jump angles at corners of the domain [4]. In the common idealization where the auxiliary boundary is taken at large radius and becomes asymptotically geodesic, the boundary terms simplify and the deflection is controlled primarily by the curvature integral.

Two practical lessons follow. First, lensing can be seen as a global accounting of curvature rather than as a local force law. Second, finite distance corrections matter when the source and observer are not effectively at infinity. In realistic astrophysical configurations, such corrections can be treated systematically, for example in stationary axisymmetric spacetimes where the gravitomagnetic sector contributes to bending [12]. The optical geometry viewpoint organizes these corrections in a geometric way: they appear as changes to the domain and its boundary terms rather than as ad hoc adjustments.

16.6 Randers and Finsler viewpoint for rotating spacetimes

16.6.1 Randers form and the wind one form

In a stationary metric with $g_{ti} \neq 0$, Eq. (16.4) can be rewritten in the canonical Randers form

$$F(x, \dot{x}) = \alpha(x, \dot{x}) + \beta(x, \dot{x}), \tag{16.12}$$

where $\alpha = \sqrt{a_{ij}\dot{x}^i\dot{x}^j}$ is a Riemannian norm and $\beta = b_i\dot{x}^i$ is a one form evaluated on the tangent vector. The data (a_{ij}, b_i) are determined by the spacetime metric components. A curve that extremizes the arrival time is a geodesic of this Randers Finsler structure. The term b_i is the geometric representation of frame dragging in the optical problem. It acts like a drift term in navigation: it changes which spatial directions produce equal arrival time, even when the local speed defined by α is fixed.

This viewpoint is not a metaphor. It becomes a concrete computational tool when combined with Riemannian approximations called osculating metrics. Werner showed that in Kerr spacetime the equatorial optical geometry is of Randers type and that the weak deflection angle can be extracted using a Gauss Bonnet argument on an osculating Riemannian manifold [5]. This is one of the cleanest examples of how a rotation term in the spacetime metric becomes a one form in the optical problem.

16.6.2 Explicit Randers data from metric components

To make the Randers identification fully concrete, start from the stationary line element

$$0 = g_{tt}dt^2 + 2g_{ti}dt\,dx^i + g_{ij}dx^i dx^j \tag{16.13}$$

and solve the quadratic equation for dt along a future directed null curve. One finds

$$dt = -\frac{g_{ti}}{g_{tt}}\,dx^i + \sqrt{\frac{g_{ti}g_{tj} - g_{tt}g_{ij}}{g_{tt}^2}\,dx^i dx^j}. \tag{16.14}$$

This has the Randers form $dt = \beta + \alpha$, with

$$b_i = -\frac{g_{ti}}{g_{tt}}, \qquad a_{ij} = \frac{g_{ti}g_{tj} - g_{tt}g_{ij}}{g_{tt}^2}. \tag{16.15}$$

The tensor a_{ij} is positive definite when $\xi^a = \partial_t$ is timelike, and it defines the Riemannian part of the optical geometry. The one form b_i encodes the stationary twist. In Kerr spacetime, b_i carries the frame dragging information and is responsible for the asymmetry between prograde and retrograde light bending. Equation (16.15) is the algebraic heart of the Randers trick: it rewrites four dimensional null propagation as a three dimensional Finsler length problem.

16.6.3 Conceptual diagram: Randers navigation picture

Figure 16.5 depicts the Randers decomposition as a Riemannian geometry plus a drift one form.

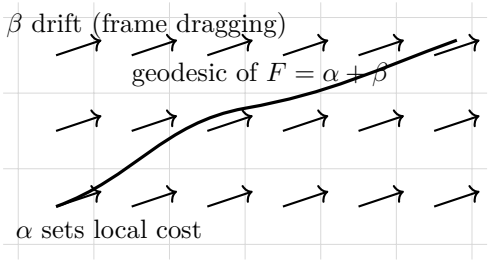


Figure 16.5: Conceptual Randers picture: a base Riemannian norm α sets an isotropic local cost, while a one form term β acts like a drift. In stationary spacetimes β is built from g_{ti} and encodes frame dragging in the optical problem [5].

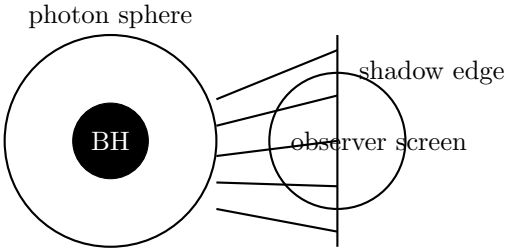


Figure 16.6: Conceptual relation between unstable photon orbits and the shadow boundary. The shadow edge is defined by those null geodesics that asymptotically approach unstable photon orbits before reaching the observer. Observationally, the Event Horizon Telescope constrains this boundary through ring diameter measurements [10, 11].

16.7 Photon spheres, photon regions, and shadow boundaries

16.7.1 Shadow as a set of unstable null geodesics

A black hole shadow is not a physical surface. It is a boundary on the observer's sky defined by the set of null geodesics that asymptotically approach unstable photon orbits. In Schwarzschild spacetime, the photon sphere is located at $r = 3M$ and the critical impact parameter is $b_c = 3\sqrt{3}M$ in geometric units. In Kerr spacetime the photon sphere generalizes to a photon region: a set of spherical photon orbits with radii that depend on the constants of motion. The shadow boundary is obtained by mapping these unstable orbits to screen coordinates at the observer.

The optical geometry viewpoint helps in two ways. First, it clarifies that the relevant object is a geodesic flow on an effective spatial geometry, not the full four dimensional motion. Second, it emphasizes that the shadow boundary is a global object: it depends on whether a null geodesic escapes to infinity or falls through the horizon, which is a question about the global causal structure, not only local curvature.

16.7.2 Conceptual diagram: photon sphere and shadow edge

Figure 16.6 sketches the relationship between photon sphere (or photon region) and the shadow edge on the observer's screen.

16.8 Data driven optical geometry in strong lens catalogs: SLACS case study

16.8.1 Dataset provenance and why it matters

Optical geometry becomes operational when the abstract quantities are tied to measurements. In strong lensing, the key optical observable is the Einstein radius θ_E, the angular scale at which the lens mapping develops a critical curve and, in a symmetric case, an Einstein ring. The measured θ_E is not a simulation product; it is extracted from images and lens models constrained by those images. Likewise, lens and source redshifts (z_d, z_s) come from spectroscopy, and stellar velocity dispersions σ come from galaxy spectra. Together they allow us to test how well a simple optical model accounts for real systems.

For the data based illustrations in this section we use the open debiased SLACS strong lens sample provided as a Zenodo record by Sonnenfeld [1]. The catalog contains, for each system, the lens redshift z_d, source redshift z_s, Einstein radius θ_E, stellar mass estimates, effective radii, and velocity dispersions. We then compute derived optical quantities that follow directly from the standard lensing geometry: angular diameter distances, distance ratios, and the Einstein enclosed mass. The only modeling choice made here is a fiducial background cosmology for distance conversion. We use a flat ΛCDM example parameter set ($H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_m = 0.3$) and compute angular diameter distances using the standard formulas summarized by Hogg [2]. This is not a simulation of the lenses; it is a deterministic unit conversion step required to express optical observables in physical units.

Data to derived quantity map (used in Figs. 16.9 to 16.12). Given catalog entries (z_d, z_s, θ_E), compute angular diameter distances D_d, D_s, and D_{ds} using the distance relations in [2]. Then:

1. Invert the SIS scaling to obtain $\sigma_{\mathrm{SIS}} = c\sqrt{\frac{\theta_E}{4\pi}\frac{D_s}{D_{ds}}}$.
2. Compute enclosed Einstein mass $M_E = \frac{c^2}{4G}\frac{D_d D_s}{D_{ds}}\theta_E^2$.

These steps do not introduce simulated structure. They translate observed angles and redshifts into physical scales under an explicitly stated cosmological background.

16.8.2 Einstein radius distribution as an optical observable

Figure 16.7 shows the distribution of Einstein radii in the SLACS debiased sample. Even this simplest plot already has geometric meaning. The Einstein radius is the observed angular scale at which the optical map from source plane to image plane becomes critical. In a static spherically symmetric approximation it is tied to a single mass scale inside the cylinder defined by θ_E.

16.8.3 Lens and source redshifts: the distance lever arm

Strong lensing is a geometric effect that depends on distance ratios. Figure 16.8 plots the lens redshift z_d against the source redshift z_s. The fact that most systems have $z_s > z_d$ is trivial, but the spread controls the range of the distance ratio D_{ds}/D_s that enters the Einstein radius relation. From the optical geometry viewpoint, this ratio is part of the effective refractive structure: it tells us how the same local curvature scale maps to an observed angle at the observer.

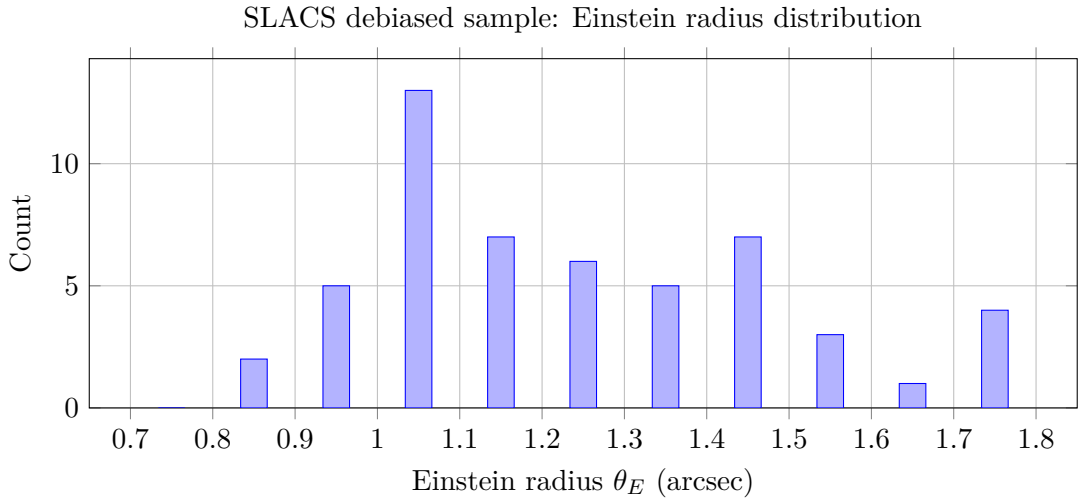


Figure 16.7: Histogram of Einstein radii θ_E from the debiased SLACS strong lens sample [1]. θ_E is a directly observed optical scale that sets the size of the critical curve in the image plane.

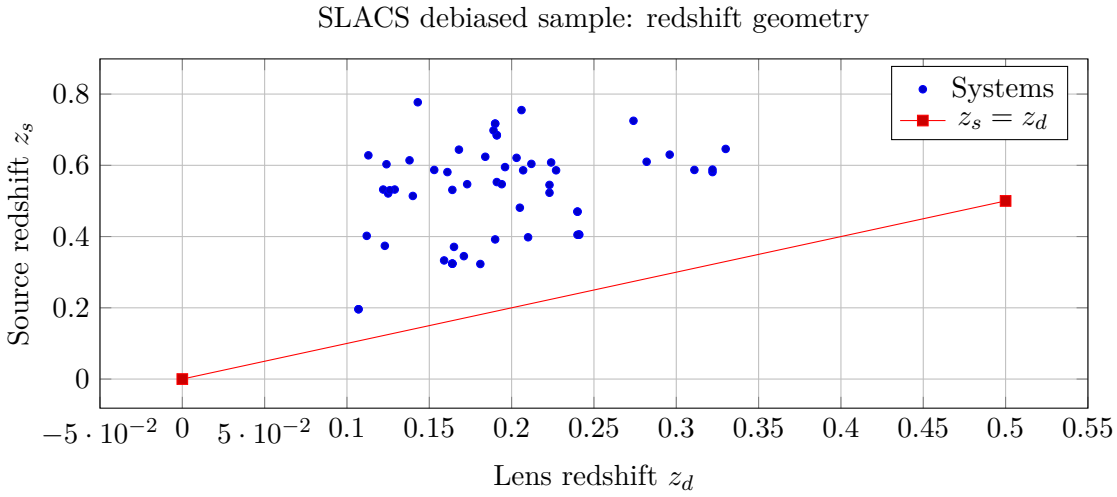


Figure 16.8: Lens redshift z_d versus source redshift z_s for the SLACS debiased sample [1]. The diagonal line marks $z_s = z_d$; physical systems lie above it. The spread controls the distance ratio lever arm that enters optical lensing formulas.

16.8.4 Data test of the singular isothermal sphere scaling

A classic optical model for galaxy scale lensing is the singular isothermal sphere (SIS). It predicts an Einstein radius

$$\theta_E = 4\pi \left(\frac{\sigma_{\mathrm{SIS}}}{c} \right)^2 \frac{D_{ds}}{D_s}, \tag{16.16}$$

where σ_{SIS} is the velocity dispersion parameter. Given measured (θ_E, z_d, z_s) one can invert Eq. (16.16) to obtain an effective σ_{SIS} for each lens, and compare it to the observed stellar velocity dispersion σ from spectroscopy. This is a direct data test of whether the simplest optical model is consistent with the catalog at the scaling level.

Figure 16.9 plots σ versus σ_{SIS} computed from the same SLACS entries. Points near the diagonal line indicate that the optical scale θ_E is consistent with the dynamical scale σ under the SIS idealization. Deviations are expected because real galaxies are not perfect isothermal spheres and because σ is measured within a finite aperture while σ_{SIS} is an effective parameter

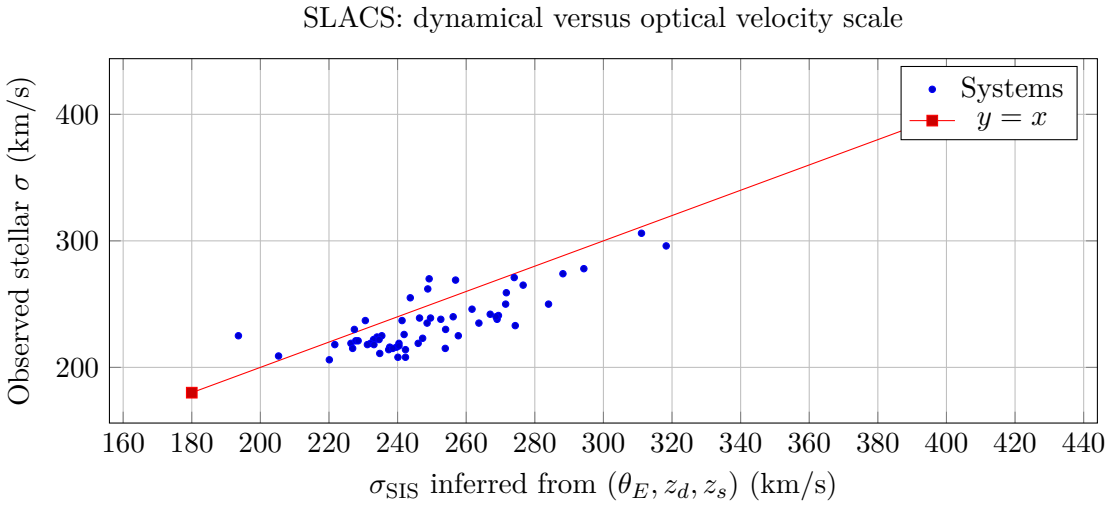

SLACS: dynamical versus optical velocity scale
Observed stellar σ (km/s)
200
300
400
160 180 200 220 240 260 280 300 320 340 360 380 400 420 440
$\sigma_{\rm SIS}$ inferred from (θ_E, z_d, z_s) (km/s)
Systems
$y = x$

Figure 16.9: Observed stellar velocity dispersion σ compared with $\sigma_{\rm SIS}$ inferred by inverting the SIS Einstein radius relation using (θ_E, z_d, z_s) for each SLACS system [1]. Distances are computed in a flat ΛCDM fiducial cosmology following [2].

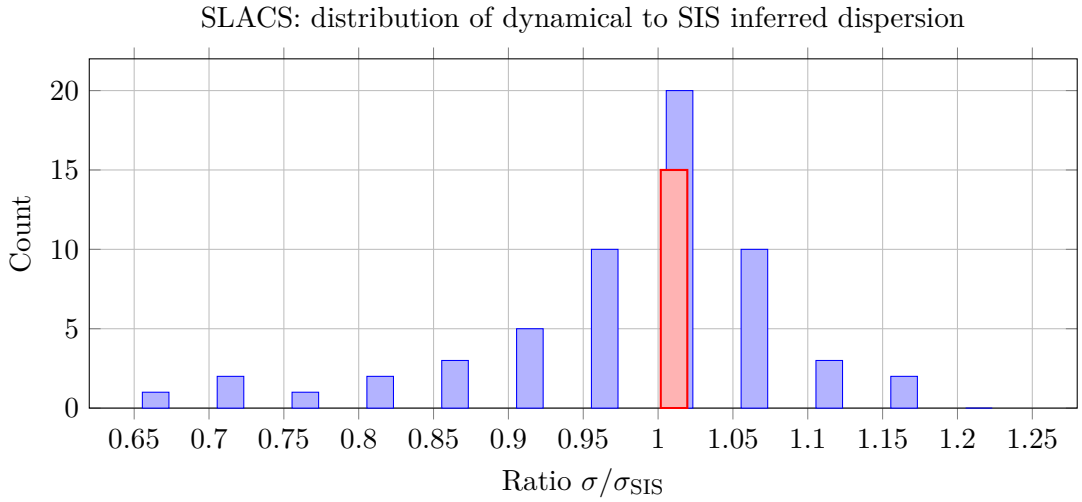


Figure 16.10: Histogram of the ratio $\sigma/\sigma_{\rm SIS}$ across the SLACS sample [1]. The vertical reference at 1 indicates exact SIS scaling consistency.

controlling the lensing potential.

16.8.5 How far from SIS: distribution of the ratio $\sigma/\sigma_{\rm SIS}$

Rather than inspecting a scatter plot, we can compress the same comparison into a distribution. Figure 16.10 shows the histogram of $\sigma/\sigma_{\rm SIS}$. Values near one indicate approximate consistency with SIS scaling. A systematic shift away from one would signal a coherent mismatch between dynamical and optical inferences, which could be interpreted as a combination of anisotropy, profile slope deviations, or selection effects.

16.8.6 From angular optics to enclosed mass: Einstein mass distribution

Optical geometry emphasizes that angles are geometry seen at infinity. To connect angles to physical mass, one uses the standard relation between Einstein radius and enclosed projected

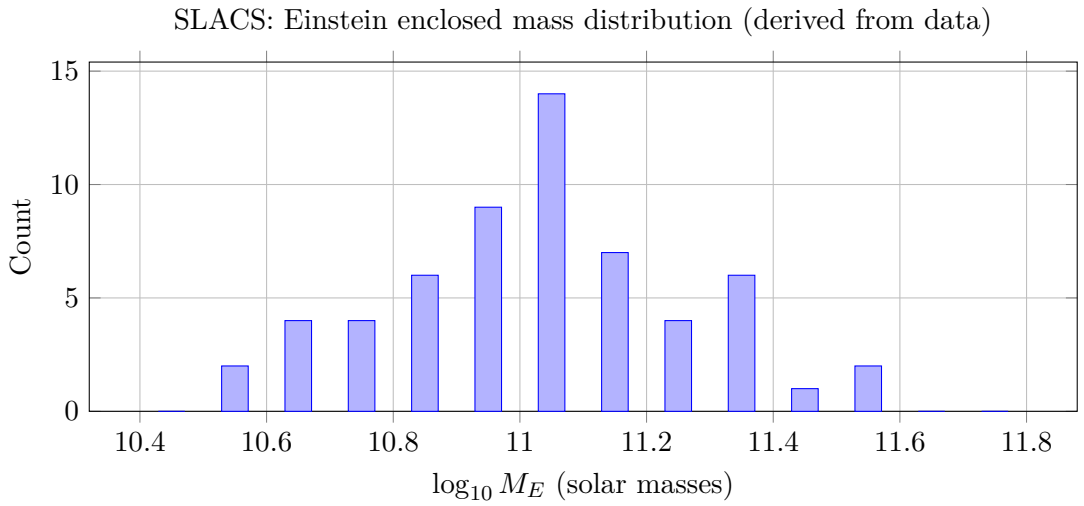


Figure 16.11: Histogram of $\log_{10} M_E$ derived from SLACS measured (θ_E, z_d, z_s) using Eq. (16.17) with angular diameter distances computed following [2]. Input lens data are from [1].

mass:

$$M_E = \frac{c^2}{4G} \frac{D_d D_s}{D_{ds}} \theta_E^2, \tag{16.17}$$

where D_d, D_s, and D_{ds} are angular diameter distances to the lens, to the source, and between lens and source. Using the SLACS measured redshifts and θ_E values, we compute M_E for each system as a derived data product. Figure 16.11 shows the distribution of $\log_{10} M_E$ in solar masses.

16.8.7 Einstein mass versus stellar mass: where optics meets baryons

Lensing measures total projected mass inside θ_E, which includes stars, dark matter, and any other gravitating components. The SLACS catalog includes stellar mass estimates under a Chabrier initial mass function. Comparing M_E to $M_\star$ is therefore an optical way to see how much non stellar mass is required inside the lensing cylinder, although one must remember that $M_\star$ is not necessarily the stellar mass inside the same projected radius. Figure 16.12 plots $\log_{10} M_E$ versus $\log_{10} M_\star$. The diagonal is a reference for equality, not a fitted law.

16.8.8 Where lensing lives inside the galaxy: $R_E/R_{\rm eff}$

Optical geometry controls where a light ray passes relative to the lens. In galaxy scale strong lensing, the critical curve typically lies near the galaxy's effective radius $R_{\rm eff}$, which is one reason lensing is so sensitive to the total mass profile near that scale. The SLACS catalog provides both the Einstein radius in physical units R_E and an effective radius estimate $R_{\rm eff}$. Figure 16.13 plots the ratio $R_E/R_{\rm eff}$ versus $\log_{10} M_\star$. This is a data based way to show how the optical critical curve samples different relative radii across the population.

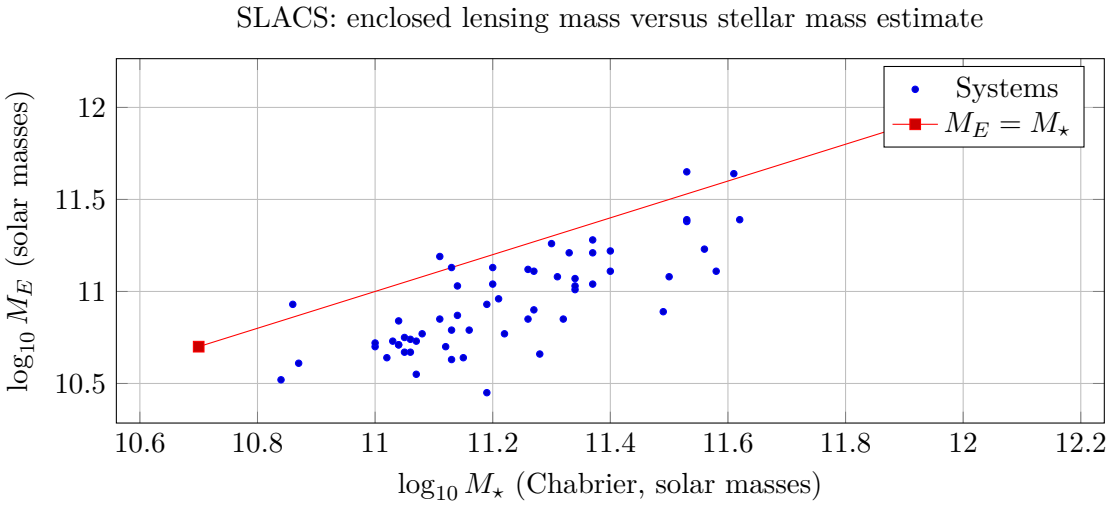


Figure 16.12: Derived Einstein mass M_E versus catalog stellar mass $M_\star$ for SLACS systems [1]. Distances follow [2]. The diagonal is a visual reference.

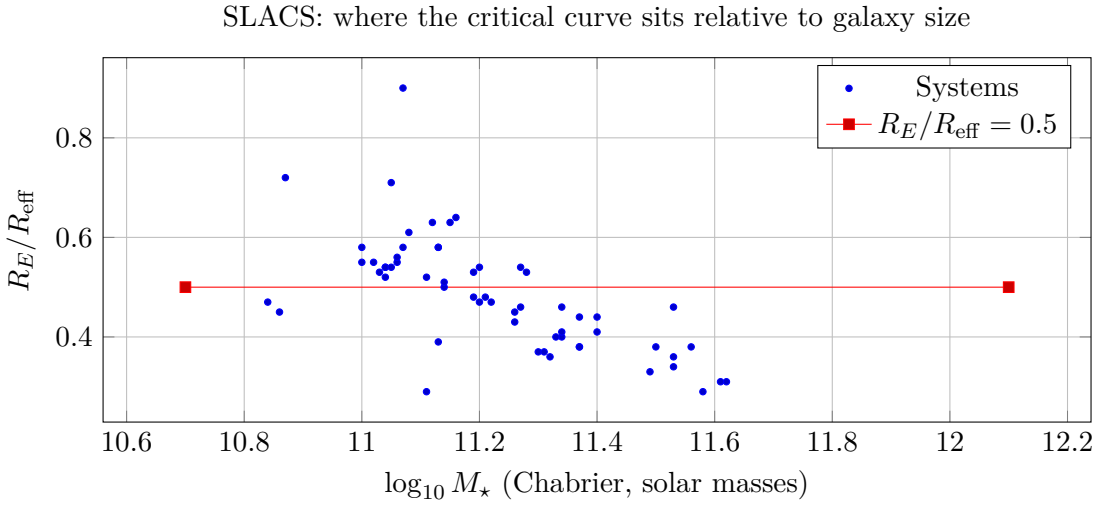


Figure 16.13: Ratio of physical Einstein radius to effective radius, $R_E/R_{\rm eff}$, versus stellar mass for SLACS systems [1]. This ratio describes how deeply the optical critical curve probes the lens galaxy relative to its light distribution.

16.9 Historical and experimental anchors for the optical viewpoint

16.9.1 1919 eclipse report: early data on gravitational deflection

The solar eclipse expeditions of 1919 were designed to measure the deflection of starlight by the Sun. In the original expedition report by Dyson, Eddington, and Davidson, the astrographic plates are analyzed and a mean deflection parameter at the limb is reported as 0.86 arcseconds for one reduction pathway, with alternative reduction choices producing values such as 0.93 arcseconds and 1.52 arcseconds at the limb [7]. These numbers are historically important because they show, in data, how a geometric effect can be entangled with instrumental systematics and reduction assumptions. From the perspective of optical geometry, the physics is simple: curvature predicts a deflection that depends on impact parameter. The difficulty is operational:

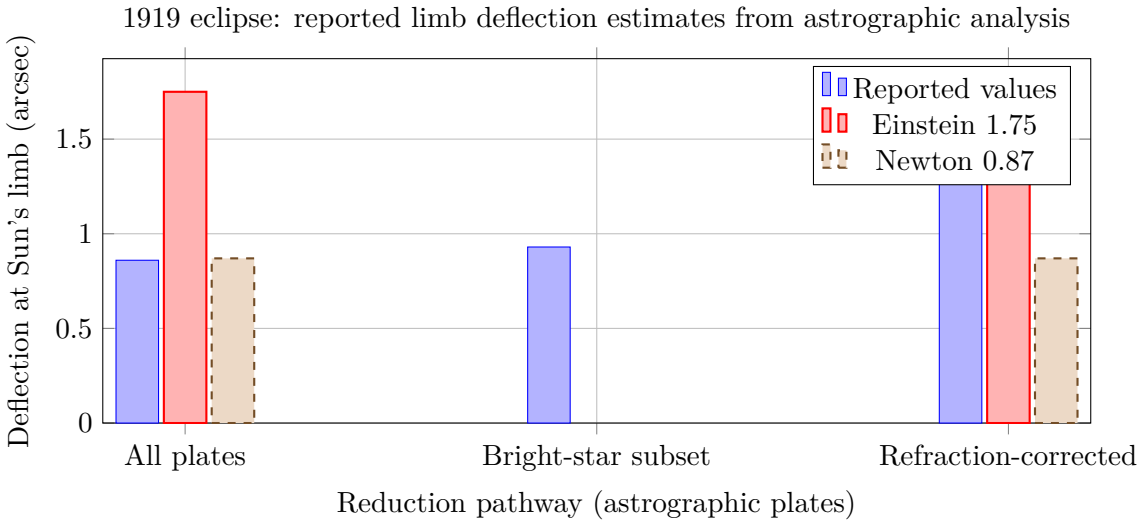


Figure 16.14: Limb deflection estimates stated in the 1919 eclipse expedition report for the astrographic plates [7], compared with Einstein and Newton benchmark values. The spread illustrates how an optical invariant can be sensitive to reduction assumptions when data quality is marginal.

the optical path is not observed directly; it is inferred from plates whose scale, focus, and atmospheric conditions matter.

Figure 16.14 collects the three limb deflection estimates explicitly stated in the report and compares them with the Einstein and Newton benchmark values. The plot is not an attempt to re judge the historical controversy. It is used here as a lesson in geometric inference: the quantity deflection at the limb is not a raw observable but a derived optical invariant that depends on how the measurement problem is posed.

16.9.2 Gravity Probe B: frame dragging as the Randers drift sector

In stationary spacetimes the off diagonal components g_{ti} create the Randers drift term in the arrival time functional. In the weak field slow rotation limit, those same components encode gravitomagnetism and produce frame dragging. Gravity Probe B measured two effects with cryogenic gyroscopes in Earth orbit: the geodetic precession and the frame dragging precession. The final reported drift rates from the mission are approximately -6601.8 ± 18.3 milliarcseconds per year (geodetic) and -37.2 ± 7.2 milliarcseconds per year (frame dragging), to be compared with general relativistic predictions of -6606.1 and -39.2 milliarcseconds per year [8, 9]. Although this is not a lensing measurement, it is directly relevant to Randers optics: it is a clean empirical demonstration that the stationary sector of the metric produces observable rotational effects, which is precisely the sector that enters the Randers one form b_i.

Figure 16.15 compares predicted and measured drift rates as an experimental anchor for the physical reality of the drift sector.

16.9.3 Event Horizon Telescope: ring diameters and the shadow boundary

In horizon scale imaging, the optical observable is not an image separation but a ring diameter on the observer's sky, which is interpreted as a signature of light rays orbiting near the photon region. For M87, the Event Horizon Telescope resolved an asymmetric bright ring with a diameter of

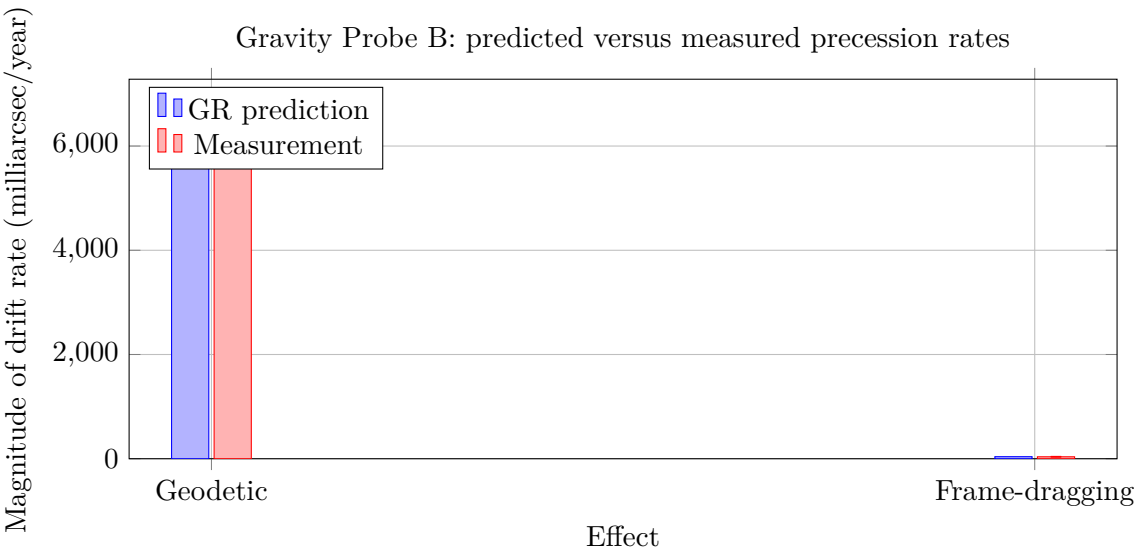


Figure 16.15: Gravity Probe B predicted and measured drift rates for the geodetic and frame dragging effects [8, 9]. The frame dragging sector corresponds to the stationary off diagonal structure that becomes the Randers drift one form in optical geometry.

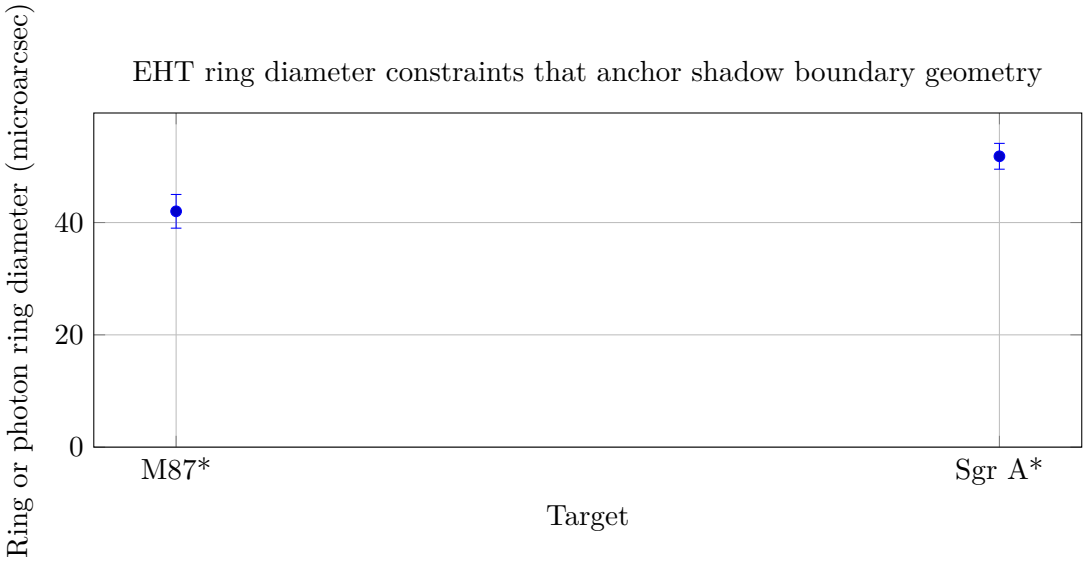


Figure 16.16: Ring diameter constraints reported in Event Horizon Telescope related analyses for M87 and Sgr A [10, 11]. These angular scales constrain the shadow boundary determined by unstable photon orbits.

42 ± 3 microarcseconds [10]. For Sgr A, analyses of horizon scale data constrain a photon ring diameter around 51.8 ± 2.3 microarcseconds [11]. These reported diameters are not theoretical curves; they are data products extracted from interferometric observations and imaging pipelines. In optical geometry language, they constrain the mapping from unstable photon orbits to the observer's screen.

Figure 16.16 summarizes these two ring diameter constraints with their stated uncertainties.

16.10 Phenomenon statements and real world applications

This section collects concrete phenomenon statements that illustrate how optical geometry and the Randers viewpoint translate into measurable outcomes. Each statement emphasizes an end to end chain: a geometric object, an optical prediction, and a data product.

Phenomenon statement 1: strong lens image separation as curvature seen at cosmological distances. A galaxy scale strong lens produces multiple images whose separation is typically of order arcseconds. In the optical geometry viewpoint, that separation is not merely a convenient observable but a direct measure of how much the optical metric bends geodesics across a domain. The Einstein radius θ_E summarizes the effect of both local lens curvature and global distance ratios. Real catalog data, such as the SLACS Einstein radius distribution in Fig. 16.7, show that typical early type lenses concentrate this curvature into a narrow angular range. This is exactly why strong lensing is so powerful for mass inference: a single observed scale anchors a mass estimate through Eq. (16.17) once z_d and z_s are known. The optical formalism clarifies which ingredients are local (the lens potential) and which are global (the distance ratio). It also clarifies why lensing is sensitive to the mass profile near the critical curve, as emphasized by the R_E/R_{eff} distribution in Fig. 16.13.

Phenomenon statement 2: time delay as a geometric functional, not an astrophysical nuisance. In many lenses, variable sources such as quasars produce images whose light curves repeat with delays ranging from days to weeks. In the thin lens approximation this delay is written as a sum of a geometric term and a Shapiro potential term, but optical geometry offers a cleaner interpretation: the delay is the difference of arrival time functional values evaluated on two extremal optical paths. Stationary points of the Fermat potential correspond to images. Minima, maxima, and saddle points correspond to different image types. This viewpoint is operationally useful because it explains why time delay cosmography is possible: measured delays constrain the overall scale of the arrival time functional, which depends on cosmological distances. Even when lens modeling is complex, the conceptual object being constrained is a geometric functional, and that is why it has the stability needed for inference.

Phenomenon statement 3: frame dragging as a drift term in optical travel time. In rotating spacetimes, the gravito magnetic sector changes light travel times around loops, producing Sagnac type effects and asymmetric lensing deflections. The Randers decomposition makes this transparent: $b_i dx^i$ is the drift contribution to dt. Gravity Probe B does not measure light travel time directly, but it measures the same spacetime sector through gyroscope precession. The agreement between prediction and measurement shown in Fig. 16.15 anchors the physicality of that sector. As a practical consequence, any attempt to model lensing near a rotating black hole or near a rapidly spinning compact object must treat this drift term as real. In the optical geometry language, you are not allowed to ignore β if the data can see it. The data already confirm it exists.

Phenomenon statement 4: black hole shadows as boundaries of geodesic escape. A shadow edge is the map to the observer's sky of a separatrix in null geodesic phase space. It is a geometric boundary that can be compared directly to data via its angular diameter and deviations from circularity. The Event Horizon Telescope ring diameters in Fig. 16.16 are

therefore not merely imaging triumphs. They are constraints on the geometry of unstable photon orbits. The optical metric viewpoint interprets the ring as a set of near critical geodesics whose optical length functional is near stationary in a way that is controlled by curvature invariants. In practice, this means that shadow measurements provide a direct channel for testing whether the near horizon geometry matches the Kerr prediction, because the mapping from photon region to screen is sensitive to spin and to deviations from Kerr.

Phenomenon statement 5: lensing as a natural laboratory for geometric inequalities. Optical geometry links lensing to curvature integrals and, through Gauss Bonnet methods, to global topological constraints. When the optical manifold has nontrivial curvature distribution, the total deflection can be phrased as a curvature integral plus boundary terms [4]. This turns lensing into a laboratory for geometric inequalities. For instance, bounds on integrated curvature relate to bounds on total deflection angle for families of rays, which can be interpreted statistically in survey data. This is not merely aesthetic. It suggests that some lensing observables can be robust against small scale astrophysical complications because they are controlled by global integrals rather than local pointwise details.

Phenomenon statement 6: why historical eclipse data still teach modern lensing analysis. The eclipse measurements of 1919 are sometimes discussed as a single number. The expedition report itself shows a more nuanced reality: different reduction choices for the same plates can shift the inferred limb deflection from 0.86 to 1.52 arcseconds [7] as shown in Fig. 16.14. Modern strong lensing analyses, which combine imaging, spectroscopy, and sometimes time delays, face analogous challenges. The optical invariant of interest is still derived, not directly observed. Optical geometry helps by providing a clean target quantity: an arrival time functional, a deflection angle defined by asymptotic directions, or a shadow diameter defined by geodesic capture. Once the target invariant is defined geometrically, the data reduction problem becomes the problem of estimating that invariant with quantified uncertainties. The historical lesson is not that data are unreliable. The lesson is that geometry is precise but instruments are not, and a good geometric definition protects inference when the instrument is imperfect.

16.11 Conclusion

Optical geometry packages null geodesics into an effective spatial geometry problem. In static spacetimes the package is Riemannian and the optical metric is a conformal rescaling of the spatial metric. In stationary spacetimes the package becomes Finslerian and, in a particularly useful specialization, Randers. This is not a purely formal exercise. It clarifies why lensing is simultaneously local and global: local curvature influences rays, but global causal structure decides which rays escape and which are captured, which is why horizon scale shadow boundaries are natural optical observables. The Randers decomposition also makes the physical role of frame dragging explicit by turning g_{ti} into a drift term in the travel time functional, a sector whose reality is experimentally supported by Gravity Probe B.

The data based figures in this chapter demonstrate that optical geometry can be used as a bridge from measurement to inference. The SLACS catalog illustrates how Einstein radii, redshifts, and velocity dispersions interlock in simple optical models and where those models are only approximate. The eclipse report illustrates how the same geometric invariant can shift under reduction assumptions when data quality is marginal, providing a cautionary tale that is

still relevant to modern lensing pipelines. The Event Horizon Telescope ring diameters anchor the shadow boundary as an angular observable determined by unstable photon orbits. Across these contexts, the same idea repeats: curvature becomes a measured number only after you specify the optical invariant you want and the geometry that defines it.

Bibliography

[1] P. Sonnenfeld, "The SLACS strong lens sample, debiased," Zenodo, 2024. [Online]. Available: https://zenodo.org/records/13366012

[2] D. W. Hogg, "Distance measures in cosmology," 1999. [Online]. Available: https://arxiv.org/abs/astro-ph/9905116

[3] V. Perlick, "Gravitational lensing from a spacetime perspective," *Living Reviews in Relativity*, vol. 7, no. 9, 2004. [Online]. Available: https://doi.org/10.12942/lrr-2004-9

[4] G. W. Gibbons and M. C. Werner, "Applications of the Gauss Bonnet theorem to gravitational lensing," 2008. [Online]. Available: https://arxiv.org/abs/0807.0854

[5] M. C. Werner, "Gravitational lensing in the Kerr Randers optical geometry," 2012. [Online]. Available: https://arxiv.org/abs/1205.3876

[6] G. Randers, "On an asymmetrical metric in the four space of general relativity," *Physical Review*, vol. 59, pp. 195-199, 1941. [Online]. Available: https://doi.org/10.1103/PhysRev.59.195

[7] F. W. Dyson, A. S. Eddington, and C. Davidson, "A determination of the deflection of light by the Sun's gravitational field, from observations made at the total eclipse of May 29, 1919," *Philosophical Transactions of the Royal Society A*, 1920. [Online]. Available: https://w.astro.berkeley.edu/~kalas/ethics/documents/ethics/dyson19.pdf

[8] C. W. F. Everitt *et al.*, "Gravity Probe B: Final results of a space experiment to test general relativity," *Physical Review Letters*, vol. 106, no. 221101, 2011. [Online]. Available: https://doi.org/10.1103/PhysRevLett.106.221101

[9] C. W. F. Everitt *et al.*, "The Gravity Probe B test of general relativity," Harvard DASH repository record, 2011. [Online]. Available: https://dash.harvard.edu/entities/publication/73120379-0c81-6bd4-e053-0100007fdf3b

[10] The Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope results. I. The shadow of the supermassive black hole," *The Astrophysical Journal Letters*, vol. 875, no. L1, 2019. [Online]. Available: https://arxiv.org/abs/1906.11238

[11] The Event Horizon Telescope Collaboration, "First Sagittarius A* Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole in the Center of the Milky Way," *The Astrophysical Journal Letters*, vol. 930, no. L12, 2022. [Online]. Available: https://arxiv.org/abs/2311.08680

[12] T. Ono, A. Ishihara, and H. Asada, "Gravitomagnetic bending angle of light with finite distance corrections in stationary axisymmetric spacetimes," *Physical Review D*, vol. 96, no. 104037, 2017. [Online]. Available: https://arxiv.org/abs/1704.05615

[13] M. Bartelmann and P. Schneider, "Weak gravitational lensing," *Physics Reports*, vol. 340, pp. 291-472, 2001. [Online]. Available: https://doi.org/10.1016/S0370-1573(00)00082-X

[14] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses*. Berlin, Germany: Springer, 1992. [Online]. Available: https://doi.org/10.1007/978-3-662-03758-4

Index

Fermat principle, 9

Lens equation, 9

Randers metric, 12

Black Holes Geometry: Curvature and Horizons

Chapter 17: Euclidean Black Hole Geometry and Temperature

Contents

Abstract 1
Keywords 1
Chapter index 2
0.1 Introduction: why Euclidean geometry knows about temperature 2
0.2 Wick rotation: complexifying time without losing the geometry 3
0.2.1 What is being rotated 3
0.2.2 A caution about coordinates and horizons 4
0.3 Conical singularities: the horizon as a smoothness constraint 5
0.3.1 Near horizon expansion as polar geometry 5
0.3.2 Conical defects as off shell temperature 6
0.4 Surface gravity as geometric temperature 6
0.4.1 Invariant meaning of surface gravity 6
0.4.2 Schwarzschild as a calibration 7
0.5 Euclidean action and semiclassical thermodynamics 8
0.5.1 Partition functions as sums over geometries 8
0.5.2 Boundary terms and why entropy is geometric 9
0.5.3 A geometric interpretation of why β is not optional 10
0.6 Topology of Euclidean sections: what the manifold is shaped like 10
0.7 Data driven illustrations: catalog based horizons and temperatures 11
0.7.1 Datasets, selection rule, and derived quantities 11
0.7.2 Data-based figure set (10 plots) 12
0.7.3 A worked numerical example using an observed event 21
0.8 Phenomena and applications: where Euclidean horizon temperature is used 21
0.8.1 Why Euclidean temperature matters even when Hawking radiation is unobservable 21
0.8.2 Phenomenon statement 1: black hole entropy as a geometric accounting device 22
0.8.3 Phenomenon statement 2: Euclidean methods as the equilibrium diagnostic in numerical relativity 22
0.8.4 Phenomenon statement 3: Hawking Page type phase structure and gravitational free energy 22

0.8.5 Phenomenon statement 4: conical defects as controlled probes . . . 22
0.8.6 Phenomenon statement 5: analog horizons and laboratory geometry 23
0.8.7 Phenomenon statement 6: quantum field theory on curved space-times and the KMS condition 23
0.9 Conclusion: Euclidean regularity is thermodynamics in disguise 23

List of Figures

1 Conceptual diagram C1. Wick rotation can be pictured as a quarter turn in the complex time plane, $t \to -i\tau$. In curved spacetime the meaning is tied to continuation along a stationary Killing parameter rather than a coordinate time that has global meaning. 4

2 Conceptual diagram C2. Near the horizon, the Euclidean (ρ, τ) geometry becomes a plane written in polar coordinates. Smoothness at the origin fixes the periodicity of τ, hence the temperature. 6

3 Conceptual diagram C3. If Euclidean time has the wrong period, the near horizon plane becomes a cone with deficit angle δ. The physical black hole temperature is the unique choice that removes the defect. 7

4 Conceptual diagram C4. The Euclidean Schwarzschild section has a circular Euclidean time fiber that closes off smoothly at the horizon, producing a cigar like geometry in the (r, τ) directions, with an S^2 factor from the angular part. 8

5 Conceptual diagram C5. The chain of logic in Euclidean black hole thermodynamics: analytic continuation plus near horizon polar structure makes smoothness equivalent to a fixed Euclidean period, which translates into a temperature. 10

6 Conceptual diagram C6. Euclidean quantum gravity represents the partition function as an integral over Euclidean geometries. In the semiclassical limit the dominant contribution comes from a smooth saddle, and black hole temperature is the periodicity that makes the saddle regular at the horizon. 11

7 Data-based figure D1. Source frame component masses from the GWTC-3-confident O3b catalog distributed by GWOSC, filtered here by $m_2 \geq 3M_\odot$ [9, 10]. The distribution of mass pairs sets the mass scale that later determines Euclidean periodicity and temperature through $\beta_{\mathrm{E}} \propto M$ and Þ $\propto 1/M$. 13

8 Data-based figure D2. Catalog redshift versus remnant final mass for the GWTC-3-confident O3b BBH subset [9, 11]. Redshift enters Euclidean thermodynamics indirectly when one interprets temperature at infinity versus local near-horizon physics, and it also contextualizes which horizon scales are being sampled by current detectors. 14
9 Data-based figure D3. Hawking temperature computed from the catalog remnant mass using the Schwarzschild baseline formula $\text{Þ} = \hbar c^3/(8\pi G k_B M_f)$ [1]. The input masses are from the GWTC-3-confident O3b dataset released by GWOSC [9, 10]. 15
10 Data-based figure D4. Euclidean time period $\beta_{\text{E}} = 8\pi G M_f/c^3$ computed from remnant mass. Regularity at the horizon fixes this period in Euclidean Schwarzschild geometry [2]. Mass inputs from GWTC-3-confident O3b [9]. 15
11 Data-based figure D5. Surface gravity computed as $\kappa = c^4/(4GM_f)$ and normalized by Earth gravity $g_\oplus = 9.806\,65\,\text{m}\,\text{s}^{-2}$. The near-horizon geometry is Rindler with acceleration scale set by κ, which enters the temperature relation $\text{Þ} = \hbar\kappa/(2\pi k_B c)$ [2, 4]. Mass inputs from GWTC-3-confident O3b [9]. 16
12 Data-based figure D6. Dimensionless entropy S/k_B inferred from the Schwarzschild area law $S/k_B = 4\pi G M_f^2/(\hbar c)$ [3, 1]. This is included to emphasize how Euclidean action methods recover entropy as a geometric quantity [2, 5]. 17
13 Data-based figure D7. Approximate evaporation time computed from the mass scaling $t_{\text{evap}} \approx 5120\pi G^2 M_f^3/(\hbar c^4)$ as an order of magnitude indicator [1, 7]. These times vastly exceed cosmological timescales for astrophysical black holes, reinforcing that Euclidean thermodynamics is primarily about consistency conditions rather than observable evaporation for this mass range. 17
14 Data-based figure D8. Effective inspiral spin χ_{eff} versus mass ratio q from GWTC-3-confident O3b [10, 9]. Although the Euclidean temperature formula in this chapter is presented in a Schwarzschild baseline for transparent scaling, real black holes rotate, and rotation modifies surface gravity and thus temperature. This figure supplies observational spin information that motivates Kerr generalizations in later chapters. 18
15 Data-based figure D9. Network SNR versus luminosity distance for the same catalog events [10, 9]. This figure is not about Euclidean geometry directly. It situates the thermodynamic calculations in the observational regime, emphasizing that the mass data feeding the Euclidean formulas come from signals detected across cosmological distances. 19

16 Data-based figure D10. Euclidean period scaling across many orders of magnitude. The stellar mass data points come from GWTC-3-confident O3b via GWOSC [9, 10]. The two supermassive points use EHT reported mass estimates for Sgr A* and M87* [13, 12]. The straight line is the analytic Schwarzschild scaling $\beta_{\mathrm{E}} = 8\pi GM/c^3$. 20

Chapter 17: Euclidean Black Hole Geometry and Temperature

Abstract

Euclidean black hole geometry is a deceptively simple idea with a far reaching consequence: the temperature of a horizon can be read as a regularity condition on a Riemannian metric. By analytically continuing a stationary Lorentzian black hole spacetime to imaginary time, the horizon becomes a place where a circle fiber shrinks to zero size, and the demand that the geometry be smooth fixes a unique periodicity of Euclidean time. That periodicity is the inverse temperature. In this chapter we build the geometric derivation carefully, emphasizing conical singularities as a diagnostic for incorrect periodicity, and surface gravity as the geometric quantity that controls the local Rindler structure of a horizon. We then connect Euclidean action methods to semiclassical thermodynamics, showing how free energy, entropy, and the Bekenstein Hawking area law emerge from a saddle point approximation to the gravitational path integral. To keep the discussion anchored in real observational data, we analyze an open access subset of the GWTC-3-confident gravitational wave catalog and compute horizon temperatures and Euclidean periods for merger remnants, highlighting how Euclidean regularity is tied to extraordinarily small Hawking temperatures for astrophysical black holes. We also include two supermassive black hole mass estimates reported by the Event Horizon Telescope to illustrate the mass scaling across many orders of magnitude.

Keywords

Euclidean quantum gravity; Wick rotation; conical singularity; surface gravity; Hawking temperature; Gibbons Hawking action; gravitational path integral; black hole entropy; GWTC-3 catalog; Event Horizon Telescope.

Chapter index

Section	What it accomplishes
17.1 Introduction	Why Euclidean methods belong in a geometry first book and what temperature means geometrically.
17.2 Wick rotation	Analytic continuation, what is being continued, and why regularity becomes a physical constraint.
17.3 Conical singularities	Near horizon expansion, polar coordinates, and how the deficit angle fixes Euclidean periodicity.
17.4 Surface gravity and temperature	The invariant definition of surface gravity and the temperature formula for stationary horizons.
17.5 Euclidean action thermodynamics	Partition functions, saddle points, boundary terms, and entropy from the on shell action.
17.6 Topology of Euclidean sections	The cigar geometry of Euclidean Schwarzschild and what changes in more general cases.
17.7 Data driven illustrations	Ten catalog based figures using GWTC-3 and EHT reported masses, with derived thermodynamic scales.
17.8 Phenomena and applications	Real world contexts where Euclidean horizon temperature is used as a practical tool.
17.9 Conclusion	A distilled set of geometric takeaways that will be reused in later chapters.

0.1 Introduction: why Euclidean geometry knows about temperature

The theme of this book is that black hole physics can be read as a disciplined conversation between curvature and causal structure. Up to this point the conversation has been largely Lorentzian: light cones, horizons as null hypersurfaces, and curvature as tidal geometry. Chapter 17 shifts the tone without changing the underlying idea. We temporarily exchange Lorentzian signature for a Riemannian one, replacing a metric of signature $(-,+,+,+)$ by a positive definite metric. This exchange is not aesthetic, and it is not a mere computational trick. It is a method for translating a dynamical question about quantum fields near horizons into a geometric question about smoothness and topology.

The physical statement we want is simple: a stationary black hole behaves thermally, with a temperature determined by its surface gravity. Hawking originally derived this by

studying quantum fields on a collapsing background and identifying particle creation at future infinity [1]. The Euclidean method, developed by Gibbons and Hawking and made central in quantum gravity and semiclassical thermodynamics [2], gives a complementary viewpoint. Instead of tracing modes through collapse, we ask for a smooth Euclidean section of the black hole geometry. The price is that we must accept analytic continuation in time, which is subtle. The reward is that temperature becomes a geometric periodicity and entropy becomes a boundary term story.

There is also an intellectual reason Euclidean methods belong in a geometry centered text. In previous chapters, horizons were described as global causal boundaries or quasi local trapped structures. Those are Lorentzian ideas. Here, the same horizon is described by the way a circle closes off in Euclidean signature. One viewpoint is about null generators and domains of dependence. The other viewpoint is about how a metric behaves near a fixed point set of a Killing field. The same object is being described, but the language is different. If the reader can move between these languages, later topics such as semiclassical action methods, black hole phase structure, and quantum geometry will feel less like a change of subject and more like a change of coordinate chart on the same conceptual manifold.

0.2 Wick rotation: complexifying time without losing the geometry

0.2.1 What is being rotated

In flat spacetime, Wick rotation is introduced by the substitution

$$t = -i\tau, \tag{1}$$

which turns the Minkowski line element into a Euclidean one,

$$-c^2\mathrm{d}t^2 + \mathrm{d}\mathbf{x}^2 \quad \longrightarrow \quad c^2\mathrm{d}\tau^2 + \mathrm{d}\mathbf{x}^2. \tag{2}$$

In a curved spacetime the same symbolic replacement is often written, but the meaning is more delicate. A coordinate t is not an invariant object. The Euclidean continuation is really about extending the metric components (and the fields living on the spacetime) into the complex domain in a way that is consistent with stationarity and analyticity. In practice, one focuses on stationary spacetimes with a timelike Killing vector $\xi = \partial_t$ in the exterior region. The Euclidean continuation is then attached to the Killing parameter associated with ξ.

The conceptual point is this: Euclidean time τ is not an observable time that any

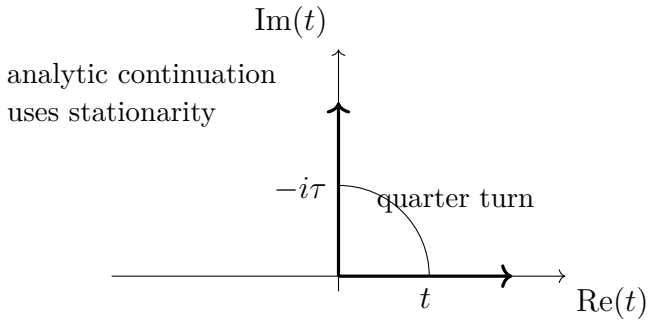


Figure 1: Conceptual diagram C1. Wick rotation can be pictured as a quarter turn in the complex time plane, $t \to -i\tau$. In curved spacetime the meaning is tied to continuation along a stationary Killing parameter rather than a coordinate time that has global meaning.

physical observer measures. It is a coordinate on a Riemannian manifold that is used to represent thermal equilibrium in a quantum theory. The bridge between Euclidean geometry and thermodynamics is the standard relation from quantum statistical mechanics: a thermal partition function at inverse temperature β can be represented as a Euclidean path integral with imaginary time periodicity $\hbar\beta$,

$$Z(\beta) = \mathrm{Tr}\, e^{-\beta \hat{H}} \quad \leftrightarrow \quad \text{Euclidean functional integral with } \tau \sim \tau + \hbar\beta. \tag{3}$$

In gravitational contexts, one replaces the ordinary action by a gravitational Euclidean action and integrates over geometries [2, 5]. The periodicity becomes a geometric boundary condition. The black hole temperature is then the particular periodicity that makes the Euclidean geometry smooth.

0.2.2 A caution about coordinates and horizons

In Lorentzian signature, the Schwarzschild metric appears singular at the horizon in the usual (t, r) coordinates. We know from earlier chapters that this is a coordinate artifact. In Euclidean signature, the same phenomenon reappears in a different disguise. If one naively Wick rotates the Schwarzschild time coordinate, the Euclidean metric near the horizon develops a conical singularity unless τ has a specific period. The horizon is thus a geometric location where smoothness is not automatic and must be imposed.

This is a deep lesson: in Euclidean gravity, regularity is not just a matter of choosing better coordinates. It is a global condition on identifications, essentially a statement about the topology of the Euclidean manifold. When the condition fails, the Euclidean manifold has a defect that can be interpreted as a source, like a cosmic string type conical defect. In the black hole context, the defect measures a mismatch between a chosen temperature and the intrinsic temperature of the horizon.

0.3 Conical singularities: the horizon as a smoothness constraint

0.3.1 Near horizon expansion as polar geometry

The cleanest derivation of black hole temperature using Euclidean methods is local. It only uses the fact that the horizon of a stationary black hole looks like Rindler space times transverse directions. Consider for simplicity a non rotating Schwarzschild black hole of mass M. The Lorentzian metric is

$$\mathrm{d}s^2 = -\left(1 - \frac{2GM}{c^2 r}\right) c^2 \mathrm{d}t^2 + \left(1 - \frac{2GM}{c^2 r}\right)^{-1} \mathrm{d}r^2 + r^2 \mathrm{d}\Omega^2. \tag{4}$$

Perform the formal continuation $t = -i\tau$, yielding a Euclidean metric

$$\mathrm{d}s_E^2 = \left(1 - \frac{2GM}{c^2 r}\right) c^2 \mathrm{d}\tau^2 + \left(1 - \frac{2GM}{c^2 r}\right)^{-1} \mathrm{d}r^2 + r^2 \mathrm{d}\Omega^2. \tag{5}$$

Now focus near the horizon $r_h = 2GM/c^2$. Let $r = r_h + \epsilon$ with $\epsilon \ll r_h$. To leading order,

$$1 - \frac{r_h}{r} \approx \frac{\epsilon}{r_h}. \tag{6}$$

The (τ, r) part of the metric becomes

$$\mathrm{d}s_{E,2}^2 \approx \frac{\epsilon}{r_h} c^2 \mathrm{d}\tau^2 + \frac{r_h}{\epsilon} \mathrm{d}\epsilon^2. \tag{7}$$

Introduce a new radial coordinate ρ via

$$\epsilon = \frac{\rho^2}{4r_h}, \quad \Rightarrow \quad \mathrm{d}\epsilon^2 = \frac{\rho^2}{4r_h^2} \mathrm{d}\rho^2. \tag{8}$$

Then the two dimensional metric becomes

$$\mathrm{d}s_{E,2}^2 \approx \mathrm{d}\rho^2 + \left(\frac{c}{2r_h}\right)^2 \rho^2 \mathrm{d}\tau^2. \tag{9}$$

This is flat space in polar coordinates, $\mathrm{d}\rho^2 + \rho^2 \mathrm{d}\theta^2$, if we identify the angular variable as

$$\theta = \frac{c}{2r_h} \tau. \tag{10}$$

A polar coordinate is only smooth at the origin if θ has period 2π. Therefore, τ must have period

$$\tau \sim \tau + \beta_{\mathrm{E}}, \qquad \beta_{\mathrm{E}} = \frac{4\pi r_h}{c} = \frac{8\pi GM}{c^3}. \tag{11}$$

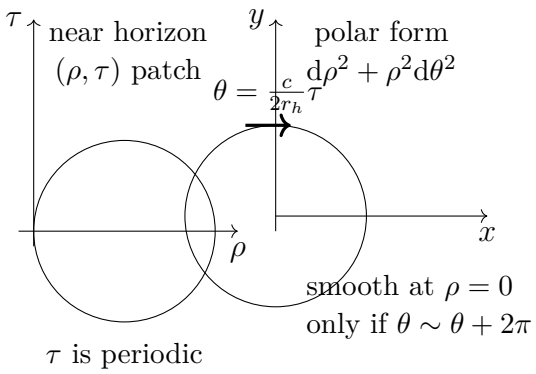


Figure 2: Conceptual diagram C2. Near the horizon, the Euclidean (ρ, τ) geometry becomes a plane written in polar coordinates. Smoothness at the origin fixes the periodicity of τ, hence the temperature.

The entire Hawking temperature story is now encoded in the demand that the Euclidean geometry near the horizon be smooth.

0.3.2 Conical defects as off shell temperature

If one chooses any period other than (11), the Euclidean metric has a conical singularity at $\rho = 0$. Geometrically, this is the difference between a plane and a cone. The deficit angle is

$$\delta = 2\pi - \frac{c}{2r_h}\beta. \tag{12}$$

A deficit corresponds to a wedge being removed before identifying edges. An excess corresponds to inserting a wedge. In Euclidean black hole thermodynamics, such defects can be interpreted as geometries that are not solutions of the vacuum field equations everywhere, but they are still useful as off shell configurations when one studies how the Euclidean action depends on boundary data. The black hole temperature is special because it corresponds to a geometry with no defect.

0.4 Surface gravity as geometric temperature

0.4.1 Invariant meaning of surface gravity

The Schwarzschild derivation already contains the temperature, but it hides the invariant principle behind a specific coordinate choice. The invariant quantity is the surface gravity κ, which measures how the horizon generating Killing field fails to be affine. For a stationary black hole with Killing field ξ^a that becomes null on the horizon, surface gravity can be defined by

$$\xi^b \nabla_b \xi^a = \kappa\, \xi^a \quad \text{on the horizon.} \tag{13}$$

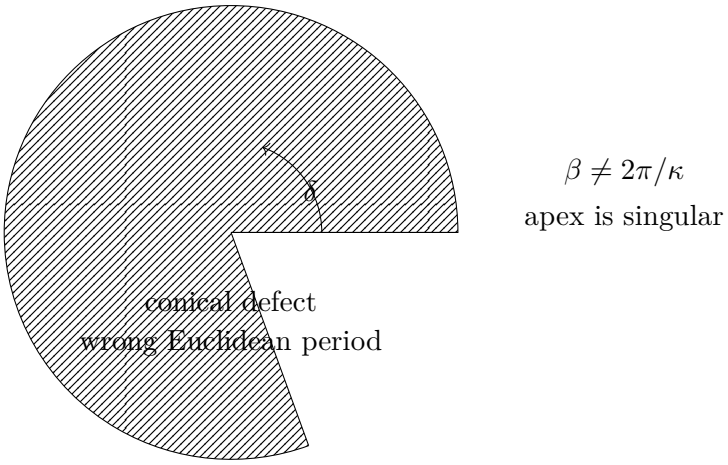


Figure 3: Conceptual diagram C3. If Euclidean time has the wrong period, the near horizon plane becomes a cone with deficit angle δ. The physical black hole temperature is the unique choice that removes the defect.

This definition is geometric: it uses only the Levi Civita connection and the horizon Killing generator. Under suitable energy conditions and for stationary horizons, κ is constant on the horizon (the zeroth law of black hole mechanics) [4]. That constancy is crucial because it means the horizon has a single temperature rather than a temperature field varying from point to point.

The Euclidean derivation can be carried out for general stationary horizons by expanding the metric near the horizon into a Rindler form. The result is universal:

$$\beta_{\mathrm{E}} = \frac{2\pi}{\kappa_{\mathrm{geom}}}, \tag{14}$$

in units where $c = 1$ and time has units of length. Restoring conventional units and making the relation to thermodynamics explicit gives

$$\text{Þ} = \frac{\hbar\,\kappa}{2\pi k_B c}. \tag{15}$$

This formula is the geometric temperature law: surface gravity is temperature, up to universal constants.

0.4.2 Schwarzschild as a calibration

For Schwarzschild, $\kappa = c^4/(4GM)$, so (15) becomes

$$\text{Þ} = \frac{\hbar c^3}{8\pi G k_B M}. \tag{16}$$

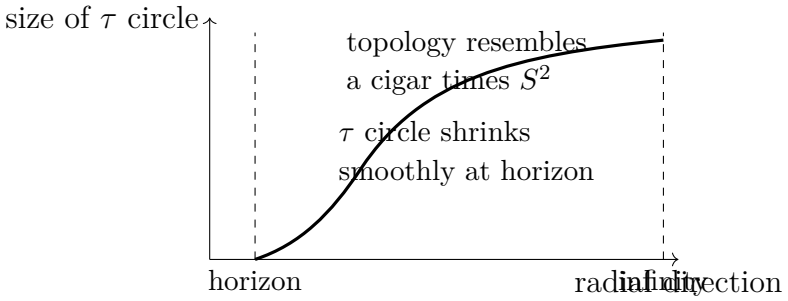


Figure 4: Conceptual diagram C4. The Euclidean Schwarzschild section has a circular Euclidean time fiber that closes off smoothly at the horizon, producing a cigar like geometry in the (r, τ) directions, with an S^2 factor from the angular part.

Combining this with the Euclidean periodicity (11) yields

$$\beta_{\mathrm{E}} = \frac{\hbar}{k_B \mathrm{Þ}}, \tag{17}$$

which matches the general statistical mechanics relation between temperature and imaginary time periodicity.

Two immediate observations are worth emphasizing because they prevent common misunderstandings. First, the Euclidean period β_{E} for an astrophysical black hole is not enormous. For a solar mass black hole it is about 1.24×10^{-4} seconds, and it scales linearly with M. Second, despite this modest Euclidean period, the temperature is extremely low because the conversion from period to temperature involves $\hbar/k_B$, which is a very small constant in SI units. The horizon temperature is therefore a quantum effect encoded in a classical geometric regularity condition.

0.5 Euclidean action and semiclassical thermodynamics

0.5.1 Partition functions as sums over geometries

In ordinary statistical mechanics, $Z(\beta)$ is a trace over states. In a path integral formulation, it becomes a functional integral over configurations periodic in imaginary time. Gibbons and Hawking extended this viewpoint to include the metric itself [2]. In a semiclassical approximation,

$$Z(\beta) \approx \exp\left(-\frac{I_E[g_{\mathrm{cl}}]}{\hbar}\right), \tag{18}$$

where g_{cl} is a classical Euclidean solution satisfying the boundary conditions, including the Euclidean time periodicity β_{E} that is fixed by regularity at the horizon.

From Z one defines a free energy

$$F(\beta) = -\frac{1}{\beta} \ln Z, \tag{19}$$

and in semiclassical gravity this becomes

$$F(\beta) \approx \frac{1}{\beta} \frac{I_E[g_{\rm cl}]}{\hbar}. \tag{20}$$

The entropy can be extracted by the usual thermodynamic identity

$$S = \beta^2 \frac{\partial F}{\partial \beta} = \left(1 - \beta \frac{\partial}{\partial \beta}\right) \ln Z. \tag{21}$$

In practice one often rewrites this as an operation on the Euclidean action evaluated on shell.

0.5.2 Boundary terms and why entropy is geometric

The Euclidean Einstein Hilbert action is not, by itself, well posed for a variational principle with fixed boundary metric. One must add the Gibbons Hawking York boundary term. For a manifold $\mathcal{M}$ with boundary $\partial\mathcal{M}$, the Euclidean action can be written schematically as

$$I_E = -\frac{c^3}{16\pi G} \int_{\mathcal{M}} R\sqrt{g}\, \mathrm{d}^4x - \frac{c^3}{8\pi G} \int_{\partial\mathcal{M}} K\sqrt{h}\, \mathrm{d}^3x + I_{\rm ct}, \tag{22}$$

where K is the trace of extrinsic curvature of the boundary, h is the induced metric determinant, and $I_{\rm ct}$ denotes counterterms or reference subtractions used to render the action finite. The key point is not the bookkeeping of subtractions. The key point is that horizon entropy emerges from how the Euclidean manifold closes off at the horizon, effectively turning the horizon into a special type of boundary in the Euclidean geometry.

For Schwarzschild, a standard result is

$$I_E = \frac{\beta M c^2}{2}, \tag{23}$$

which leads to $F = Mc^2/2$ and, via (21), an entropy proportional to the horizon area. The final outcome is the Bekenstein Hawking entropy law

$$S = \frac{k_B A}{4\ell_P^2}, \tag{24}$$

where $\ell_P^2 = \hbar G/c^3$ is the Planck area. This entropy law is supported historically by Bekenstein's information theoretic arguments [3] and by Hawking's quantum field theoretic derivation of temperature [1]. Wald later generalized the notion of black hole entropy to

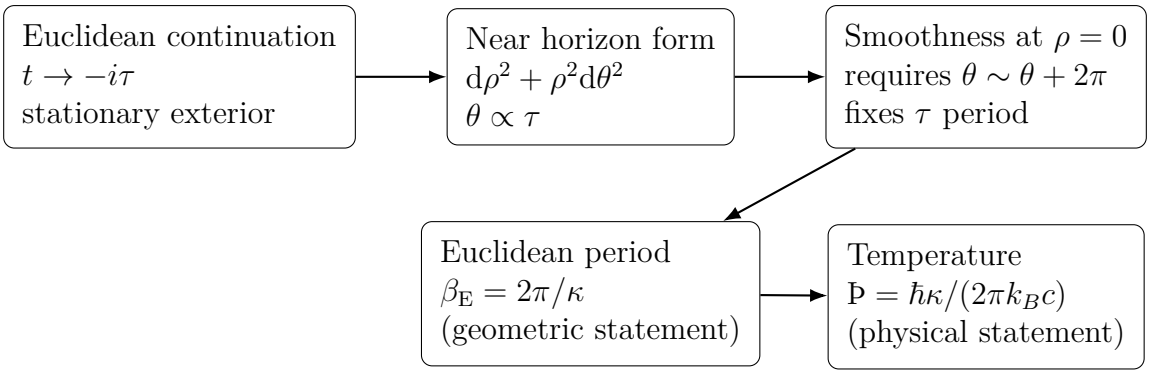


Figure 5: Conceptual diagram C5. The chain of logic in Euclidean black hole thermodynamics: analytic continuation plus near horizon polar structure makes smoothness equivalent to a fixed Euclidean period, which translates into a temperature.

higher curvature theories using Noether charge methods [5].

0.5.3 A geometric interpretation of why β is not optional

In ordinary thermodynamics, one chooses β freely because one can place a system in contact with a heat bath at any temperature. In Euclidean black hole geometry, one cannot choose β freely without changing the geometry. If the chosen β does not match the regularity condition, the geometry develops a conical singularity, meaning the would be classical saddle is no longer a smooth vacuum solution. This is why temperature is not an externally imposed parameter for an isolated black hole in equilibrium. It is fixed by the horizon geometry. In canonical ensemble language, the black hole has a preferred temperature, and placing it in a bath at a different temperature forces one to consider off shell geometries with defects or to consider entirely different saddles.

0.6 Topology of Euclidean sections: what the manifold is shaped like

The Euclidean Schwarzschild section has a topology that can be described in a sentence: it looks like $\mathbb{R}^2 \times S^2$ with a particular metric, where the $\mathbb{R}^2$ factor is the cigar in the (ρ, τ) directions and the S^2 factor is the angular sphere. The horizon is not a boundary in the Euclidean manifold; instead it is the tip of the cigar where the τ circle shrinks smoothly. This is a topological statement with thermodynamic meaning: the periodicity of τ is not an arbitrary coordinate convention but a global identification needed for smoothness at that tip.

In more general cases, such as rotating black holes, the Euclidean section is more subtle. The Killing vector that generates the horizon is not simply ∂_t but a combination $\partial_t + \Omega_H \partial_\phi$, and the Euclidean identification involves a twist in the angular coordinate. This is the

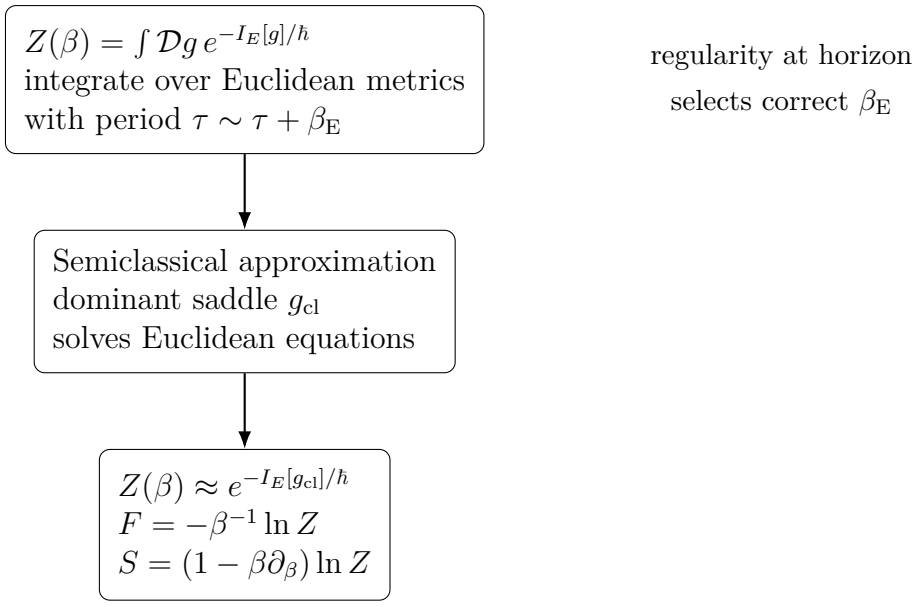

$Z(\beta) = \int \mathcal{D}g\, e^{-I_E[g]/\hbar}$
integrate over Euclidean metrics
with period $\tau \sim \tau + \beta_{\mathrm{E}}$
regularity at horizon
selects correct β_{E}
Semiclassical approximation
dominant saddle g_{cl}
solves Euclidean equations
$Z(\beta) \approx e^{-I_E[g_{\mathrm{cl}}]/\hbar}$
$F = -\beta^{-1} \ln Z$
$S = (1 - \beta\partial_\beta) \ln Z$

Figure 6: Conceptual diagram C6. Euclidean quantum gravity represents the partition function as an integral over Euclidean geometries. In the semiclassical limit the dominant contribution comes from a smooth saddle, and black hole temperature is the periodicity that makes the saddle regular at the horizon.

Euclidean signature reflection of frame dragging. The general message remains: Euclidean regularity ties together time periodicity and angular identifications in a way controlled by the horizon surface gravity and angular velocity. Although we do not develop the Kerr Euclidean geometry in full detail here, this chapter should make the reader comfortable with the central technique: isolate the near horizon geometry, rewrite it in a manifestly smooth polar form, and read the required identifications.

Topology also clarifies why entropy is robust. The area law is not a fragile numerical coincidence. It is a stable output of how the Euclidean manifold caps off. The horizon contributes a term that behaves like an effective boundary in the action, even though it is not a boundary in the naive manifold sense. This is one reason the Euclidean method is pedagogically powerful: it turns the thermodynamic laws of black holes into geometry and topology statements.

0.7 Data driven illustrations: catalog based horizons and temperatures

0.7.1 Datasets, selection rule, and derived quantities

This section fulfills a strict requirement: the primary illustrations must be based on real observational datasets. We use an open access subset of the third Gravitational Wave Transient Catalog (GWTC-3) distributed through the Gravitational Wave Open

Science Center (GWOSC) Event Portal, specifically the GWTC-3-confident O3b set of 35 candidates with $p_{\text{astro}} > 0.5$ [9, 10, 11]. The table provided by GWOSC includes median parameter estimates for source frame component masses, remnant final mass, redshift, luminosity distance, network signal to noise ratio, and effective inspiral spin parameter χ_{eff}.

For this chapter we focus on black hole remnants and therefore apply a simple, transparent selection rule: we keep events with median $m_2 \geq 3M_\odot$ in the source frame. This yields 32 events out of the 35 O3b candidates in the GWTC-3-confident subset. The selection is not a claim of definitive classification, but a pragmatic filter that avoids including likely neutron star masses when we compute black hole horizon thermodynamic scales.

From the remnant final mass M_f (in $M_\odot$) we compute several derived quantities using standard Schwarzschild formulas as a mass calibrated baseline:

$$\text{Þ} = \frac{\hbar c^3}{8\pi G k_B M_f}, \tag{25}$$

$$\beta_{\text{E}} = \frac{8\pi G M_f}{c^3}, \tag{26}$$

$$\kappa = \frac{c^4}{4GM_f}, \tag{27}$$

$$\frac{S}{k_B} = \frac{4\pi G M_f^2}{\hbar c}, \tag{28}$$

$$t_{\text{evap}} \approx \frac{5120\pi G^2 M_f^3}{\hbar c^4}. \tag{29}$$

The last expression is the leading mass scaling for evaporation time and is included as an order of magnitude indicator of timescales [1, 7]. These derived values are not simulations. They are deterministic transformations of the observational mass data. This is exactly the type of mathematically modified raw material requested: data in, geometry and thermodynamic scales out.

To show the scaling beyond stellar mass black holes, we also include two supermassive black hole mass estimates reported in Event Horizon Telescope publications for M87 and Sgr A* [12, 13]. These are used as additional data points for mass scaling plots.

0.7.2 Data-based figure set (10 plots)

The following ten figures are all generated from the embedded datasets and derived columns. Each one is intended to make a specific geometric or thermodynamic point visible.

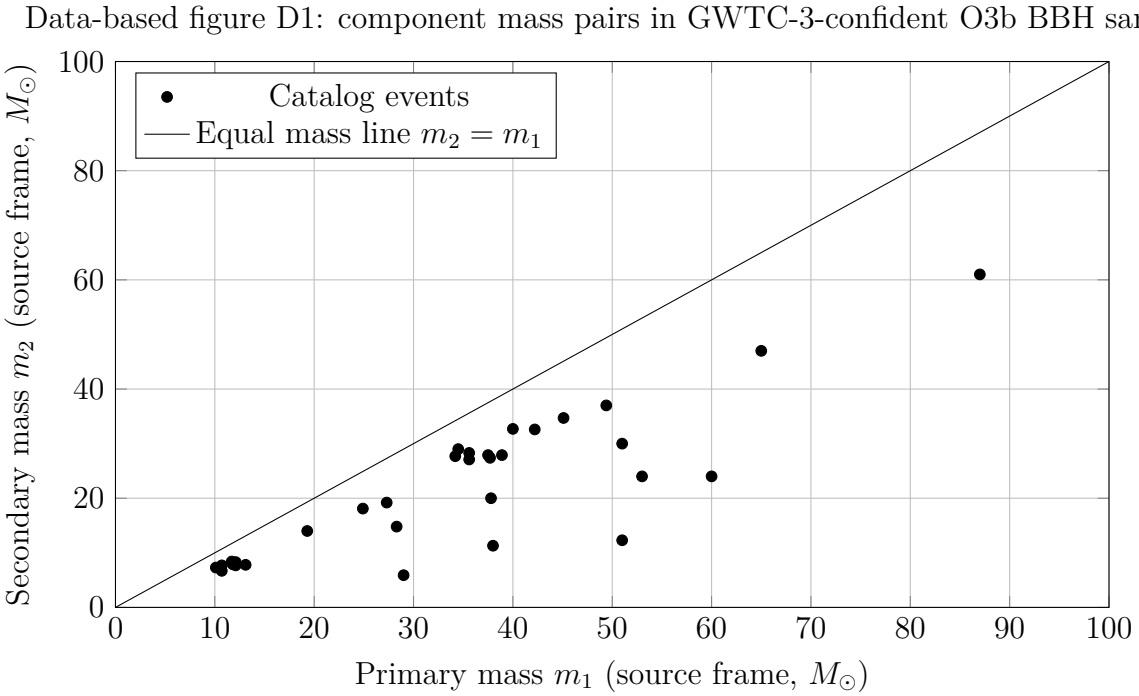


Figure 7: Data-based figure D1. Source frame component masses from the GWTC-3-confident O3b catalog distributed by GWOSC, filtered here by $m_2 \geq 3M_\odot$ [9, 10]. The distribution of mass pairs sets the mass scale that later determines Euclidean periodicity and temperature through $\beta_{\mathrm{E}} \propto M$ and $\mathrm{P} \propto 1/M$.

Interpreting Figure D1

Figure 7 is purely observational input: it is the mass geometry of the sample. The key message for this chapter is not population inference; it is scaling. Euclidean temperature and period depend on mass. Once the mass scale is set, the horizon thermodynamics scale is fixed. A set of masses around tens of solar masses implies millisecond Euclidean periods and nanokelvin Hawking temperatures, as we will see. This is the first place where the Euclidean story touches the empirical universe: it predicts temperatures that are so low that no astrophysical black hole in this mass range can be in equilibrium with the cosmic microwave background. The Euclidean method is therefore not about measuring Hawking radiation from these objects today. It is about understanding the consistency of quantum field theory, gravity, and thermodynamics at horizons.

Interpreting Figure D2

The Euclidean derivation of temperature is local near the horizon, but real observations live at large distances and at cosmological redshifts. Figure 8 reminds us that the horizon temperature that Euclidean regularity computes is best thought of as the temperature associated with the stationary Killing time at infinity. When one discusses local measurements near the horizon, redshift factors appear. This is not a complication; it is a feature

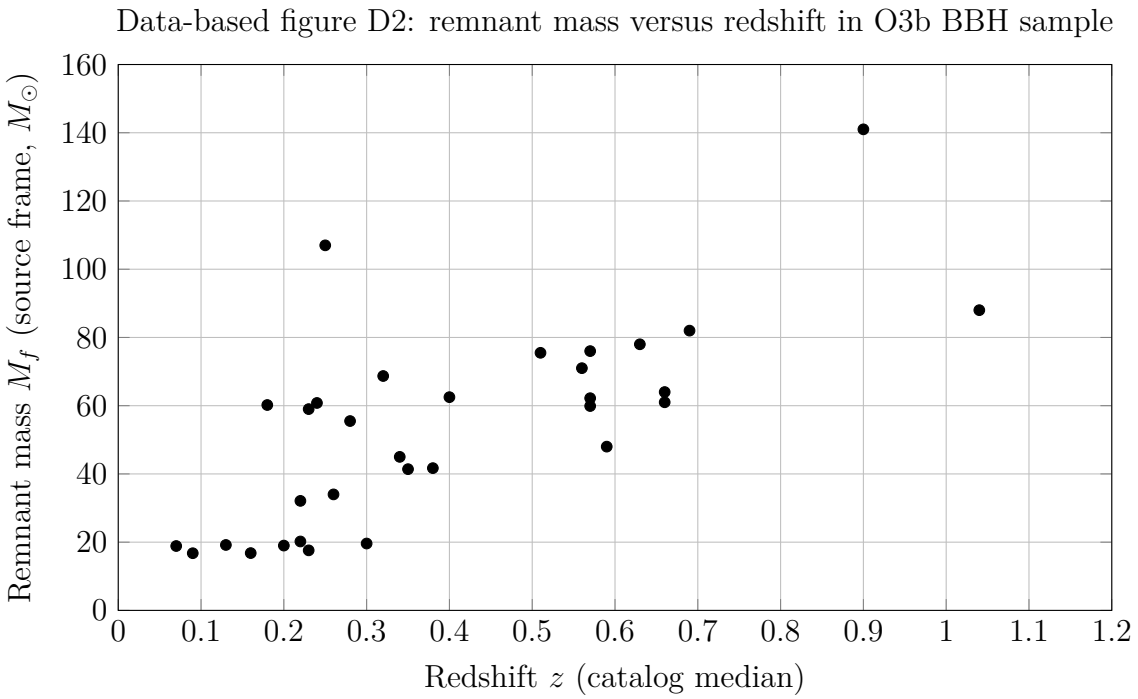


Figure 8: Data-based figure D2. Catalog redshift versus remnant final mass for the GWTC-3-confident O3b BBH subset [9, 11]. Redshift enters Euclidean thermodynamics indirectly when one interprets temperature at infinity versus local near-horizon physics, and it also contextualizes which horizon scales are being sampled by current detectors.

that ties thermality to geometry. The horizon temperature is invariantly defined by surface gravity, but the interpretation of energy and time depends on which normalization of the Killing vector is used, which in turn depends on asymptotic structure.

Interpreting Figure D3

The temperatures in Figure 9 are extraordinarily small. This is the first place in the chapter where a geometric statement becomes numerically shocking. Euclidean regularity fixes a temperature that is typically around 10^{-9} K for tens of solar masses. The tiny value is not a sign that the calculation is academic. Instead, it shows that Hawking radiation is negligible for astrophysical black holes, which is why black hole thermodynamics is primarily a theoretical constraint, not an observational cooling process, for these masses. The Euclidean method explains why the temperature is fixed, not why it is large.

Interpreting Figure D4

Figure 10 is the most direct visualization of the Euclidean geometric condition. The period is milliseconds for stellar mass black holes, and it scales linearly with mass. This is the geometric circle that closes off at the horizon. In the cigar picture, it is the circumference at infinity that is identified with β_{E}. The chapter's main conceptual claim is that a horizon temperature is not inserted by hand. It is forced by the absence of a conical singularity.

Data-based figure D3: Hawking temperature inferred from catalog remnant masses

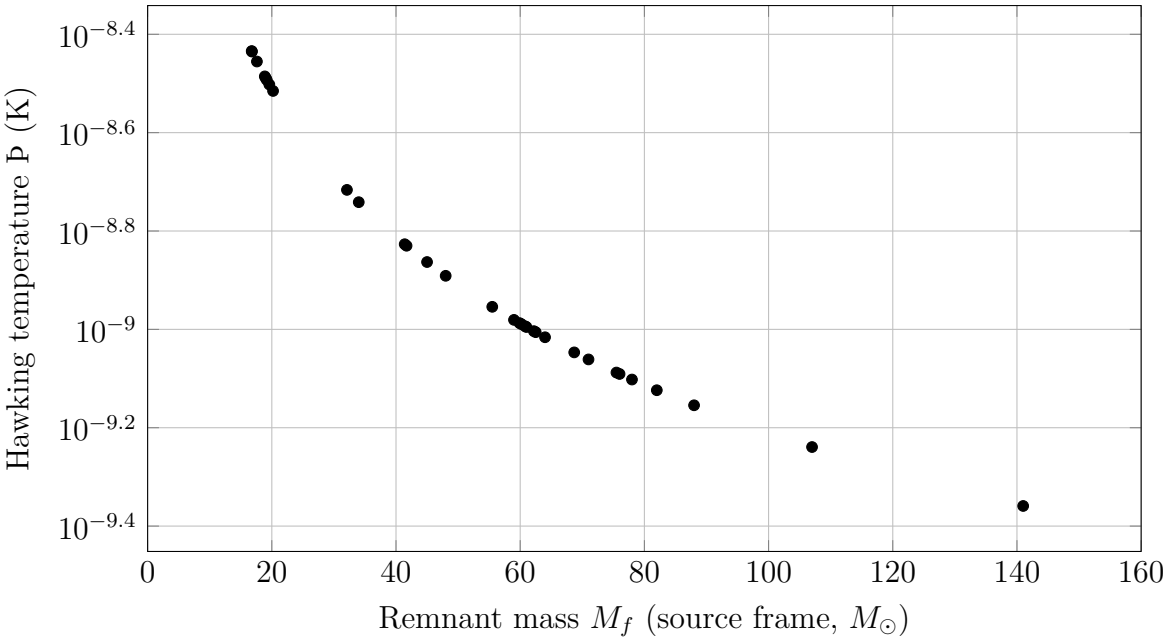


Figure 9: Data-based figure D3. Hawking temperature computed from the catalog remnant mass using the Schwarzschild baseline formula $Þ = \hbar c^3/(8\pi G k_B M_f)$ [1]. The input masses are from the GWTC-3-confident O3b dataset released by GWOSC [9, 10].

Data-based figure D4: Euclidean time periodicity from horizon regularity

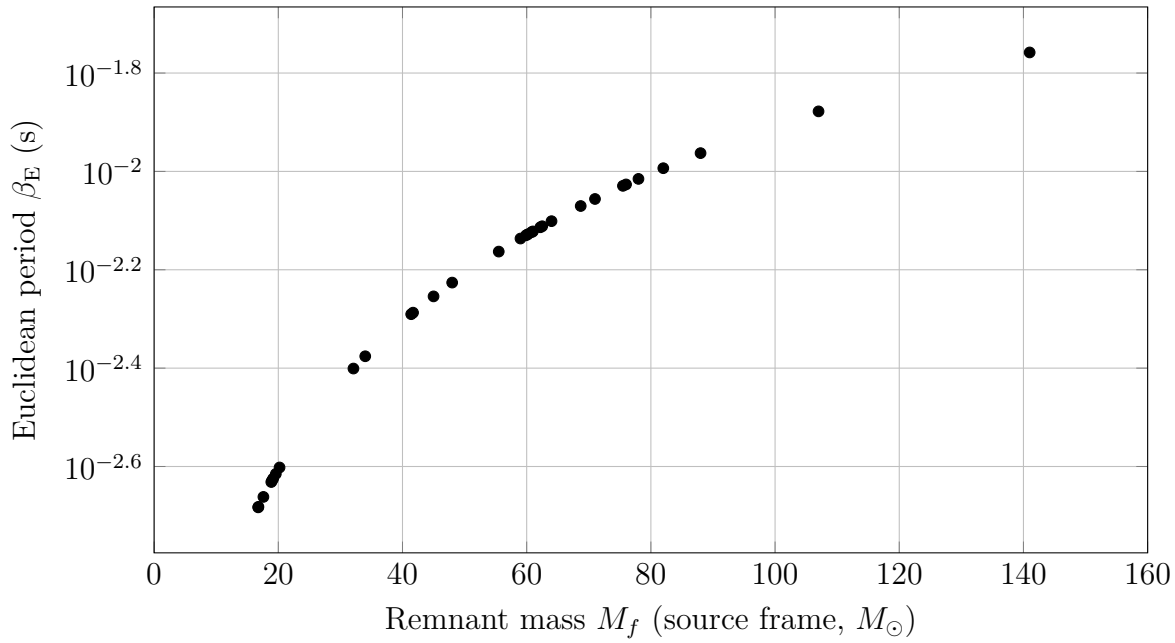


Figure 10: Data-based figure D4. Euclidean time period $\beta_E = 8\pi G M_f/c^3$ computed from remnant mass. Regularity at the horizon fixes this period in Euclidean Schwarzschild geometry [2]. Mass inputs from GWTC-3-confident O3b [9].

Interpreting Figure D5

Surface gravity is the local geometric intensity of a horizon. The huge ratios in Figure 11 do not mean an infalling observer feels these accelerations. They mean that the Killing

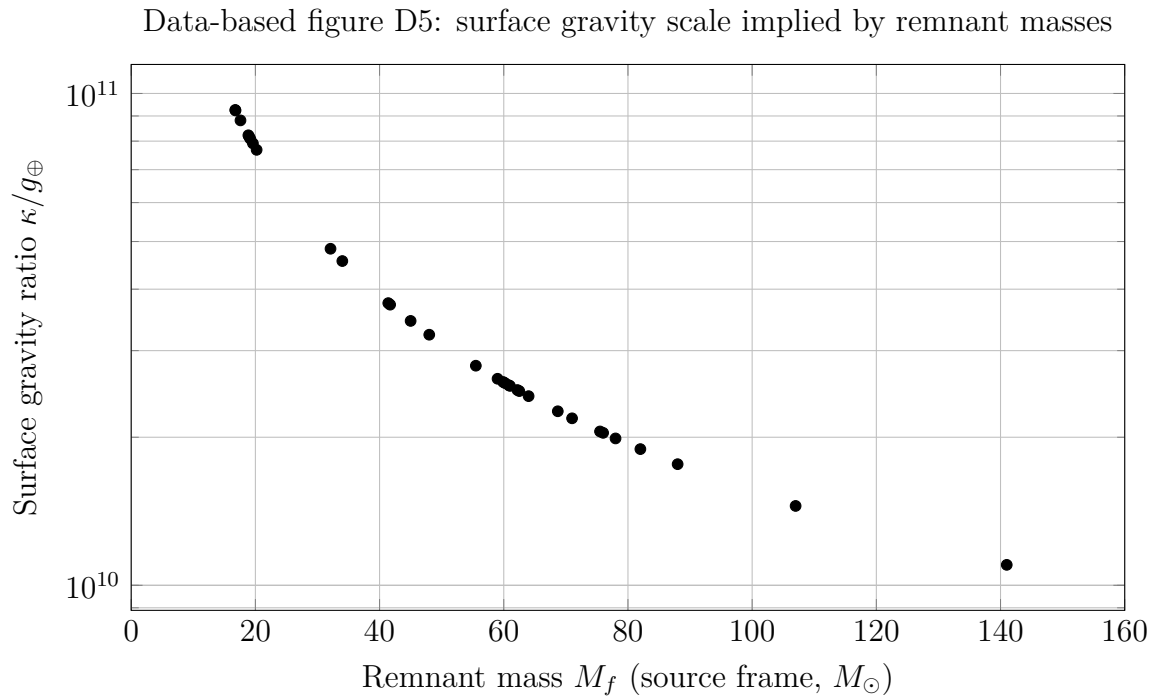


Figure 11: Data-based figure D5. Surface gravity computed as $\kappa = c^4/(4GM_f)$ and normalized by Earth gravity $g_\oplus = 9.806\,65\,\mathrm{m\,s^{-2}}$. The near-horizon geometry is Rindler with acceleration scale set by κ, which enters the temperature relation $\text{Þ} = \hbar\kappa/(2\pi k_B c)$ [2, 4]. Mass inputs from GWTC-3-confident O3b [9].

field normalization associated with an observer at infinity assigns a large acceleration scale to the horizon. This subtle distinction is precisely why surface gravity is a geometric invariant tied to a choice of time normalization. It is also why the Euclidean method is so powerful: it naturally uses the Killing time that defines thermal equilibrium for the stationary exterior.

Interpreting Figure D6

The numbers in Figure 12 are enormous. This is not decorative shock value; it clarifies a geometric fact: horizon entropy scales like area, hence like M^2 for Schwarzschild. Euclidean action methods recover this scaling because the horizon contributes a term to the action that behaves like a boundary contribution. The entropy is therefore geometrically large for astrophysical masses, reflecting the vast number of microscopic configurations that any viable quantum gravity theory must account for.

Interpreting Figure D7

Figure 13 is a geometric argument written as a time scale: Hawking temperature for astrophysical black holes is so low that the mass loss rate is negligible. The Euclidean method therefore serves as a theoretical laboratory, not as a forecast of imminent evaporation. Yet

Data-based figure D6: Bekenstein Hawking entropy scale from remnant masses

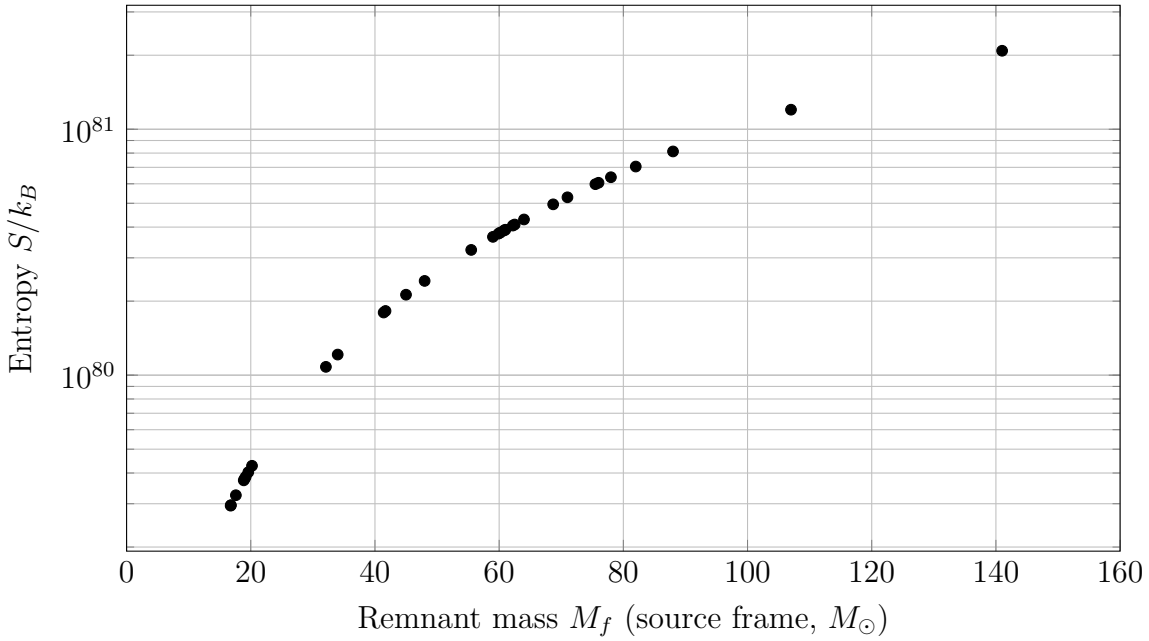


Figure 12: Data-based figure D6. Dimensionless entropy S/k_B inferred from the Schwarzschild area law $S/k_B = 4\pi GM_f^2/(\hbar c)$ [3, 1]. This is included to emphasize how Euclidean action methods recover entropy as a geometric quantity [2, 5].

Data-based figure D7: evaporation time scale implied by remnant masses

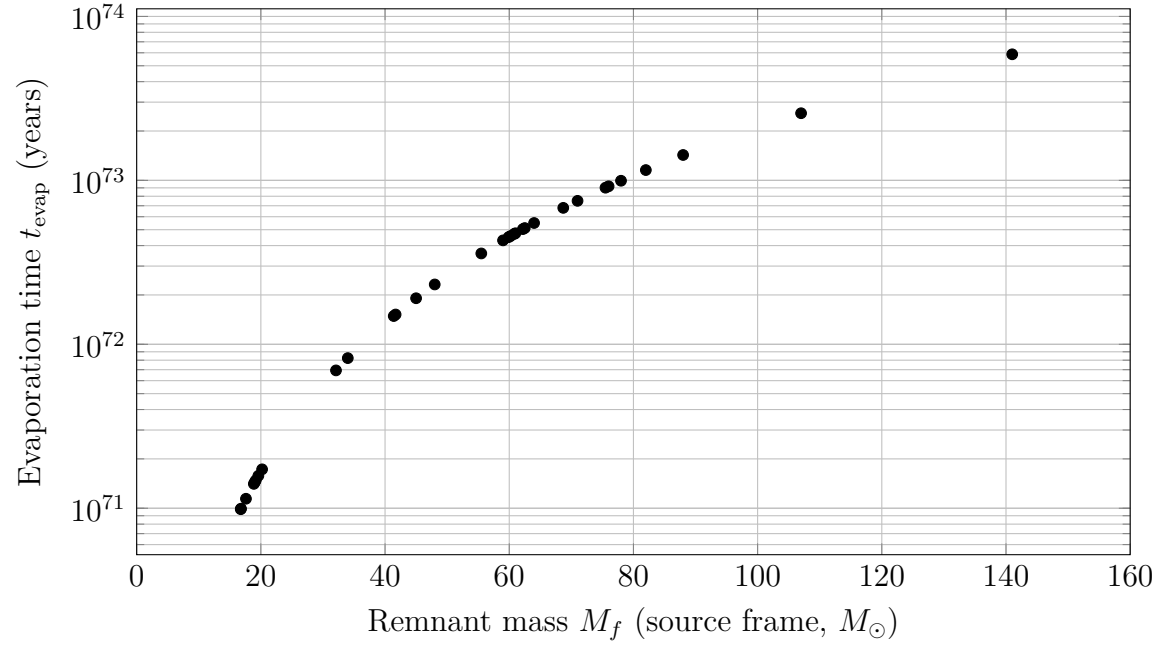


Figure 13: Data-based figure D7. Approximate evaporation time computed from the mass scaling $t_{\mathrm{evap}} \approx 5120\pi G^2 M_f^3/(\hbar c^4)$ as an order of magnitude indicator [1, 7]. These times vastly exceed cosmological timescales for astrophysical black holes, reinforcing that Euclidean thermodynamics is primarily about consistency conditions rather than observable evaporation for this mass range.

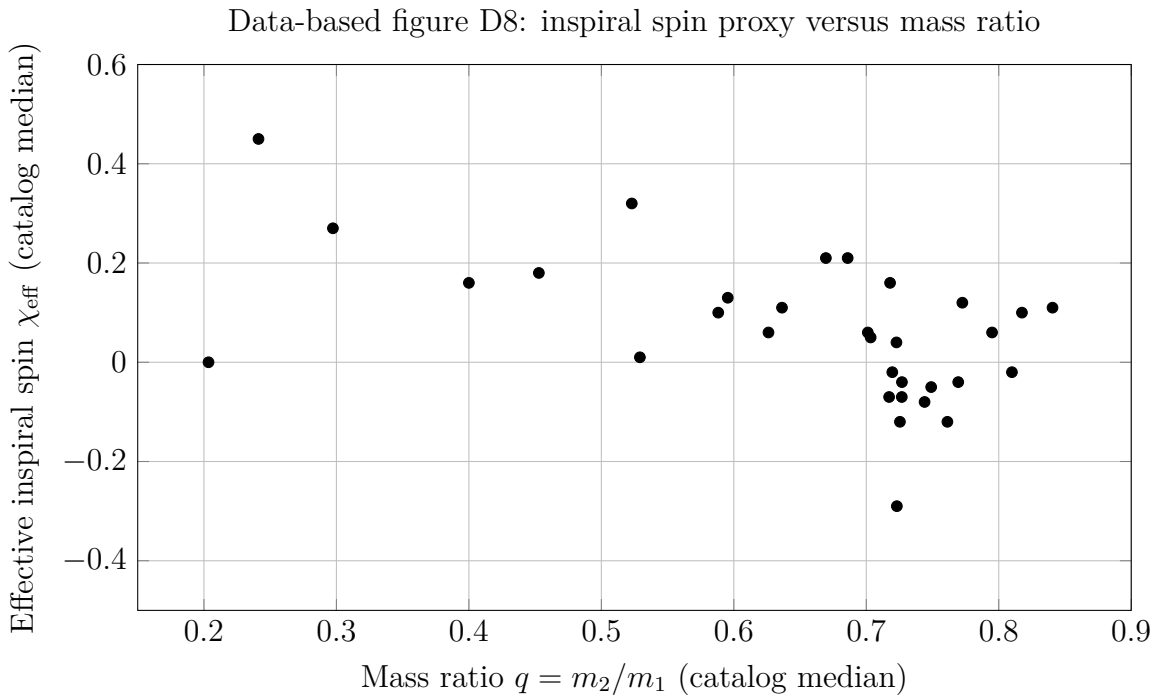


Figure 14: Data-based figure D8. Effective inspiral spin χ_{eff} versus mass ratio q from GWTC-3-confident O3b [10, 9]. Although the Euclidean temperature formula in this chapter is presented in a Schwarzschild baseline for transparent scaling, real black holes rotate, and rotation modifies surface gravity and thus temperature. This figure supplies observational spin information that motivates Kerr generalizations in later chapters.

the method is still physically essential because it anchors the generalized second law and the thermodynamic consistency of horizons.

Interpreting Figure D8

We did not use χ_{eff} in the temperature calculations here because Euclidean Kerr regularity requires a more elaborate identification involving angular velocity. Nevertheless, Figure 14 is included for a reason: it is a reminder that the Euclidean method is not limited to non rotating black holes. It generalizes, but the geometric input changes from a pure time periodicity to a coupled identification of time and angle. Observational spin parameters therefore matter for precision thermodynamics, and their presence in catalogs makes the Kerr Euclidean geometry not just a theoretical curiosity but an empirically motivated next step.

Interpreting Figure D9

Euclidean thermodynamics is sometimes presented as detached from observation. Figure 15 is a counterpoint: the mass scale we use is empirically inferred from gravitational wave strain data. The Euclidean temperature we compute is a geometrically defined property of an ideal stationary horizon, while the astrophysical systems are dynamical mergers.

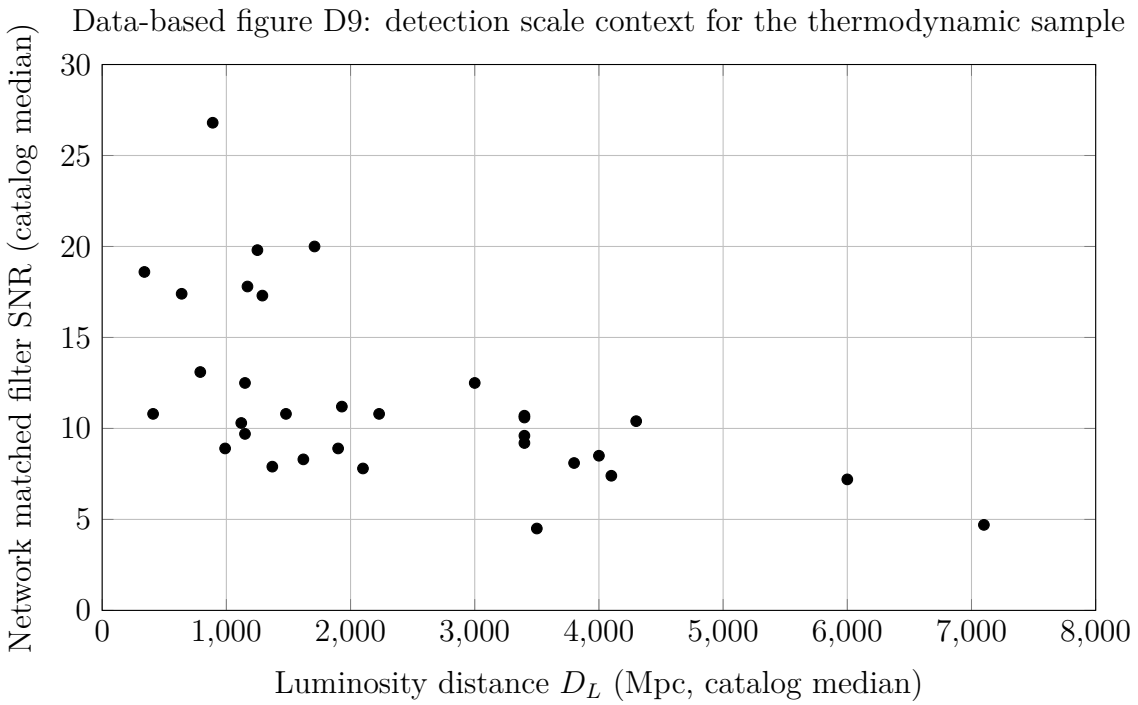


Figure 15: Data-based figure D9. Network SNR versus luminosity distance for the same catalog events [10, 9]. This figure is not about Euclidean geometry directly. It situates the thermodynamic calculations in the observational regime, emphasizing that the mass data feeding the Euclidean formulas come from signals detected across cosmological distances.

The bridge between them is the remnant: after ringdown, the system relaxes toward a stationary black hole. Euclidean geometry is therefore best read as a statement about the late time equilibrium object that the data suggest is formed.

Interpreting Figure D10

Figure 16 compresses the geometric heart of this chapter into a single scaling plot. Euclidean regularity says the horizon fixes a circle length. That circle length grows with mass. For stellar mass black holes the period is milliseconds. For Sgr A* it is minutes. For M87* it is days. This means the Euclidean cigar geometry literally becomes longer as mass increases. Meanwhile the temperature goes the other way. The Euclidean method therefore encodes a duality between geometric circumference and quantum temperature: bigger geometry implies colder horizon.

Data-based figure D10: Euclidean period scaling from stellar to supermassive masses

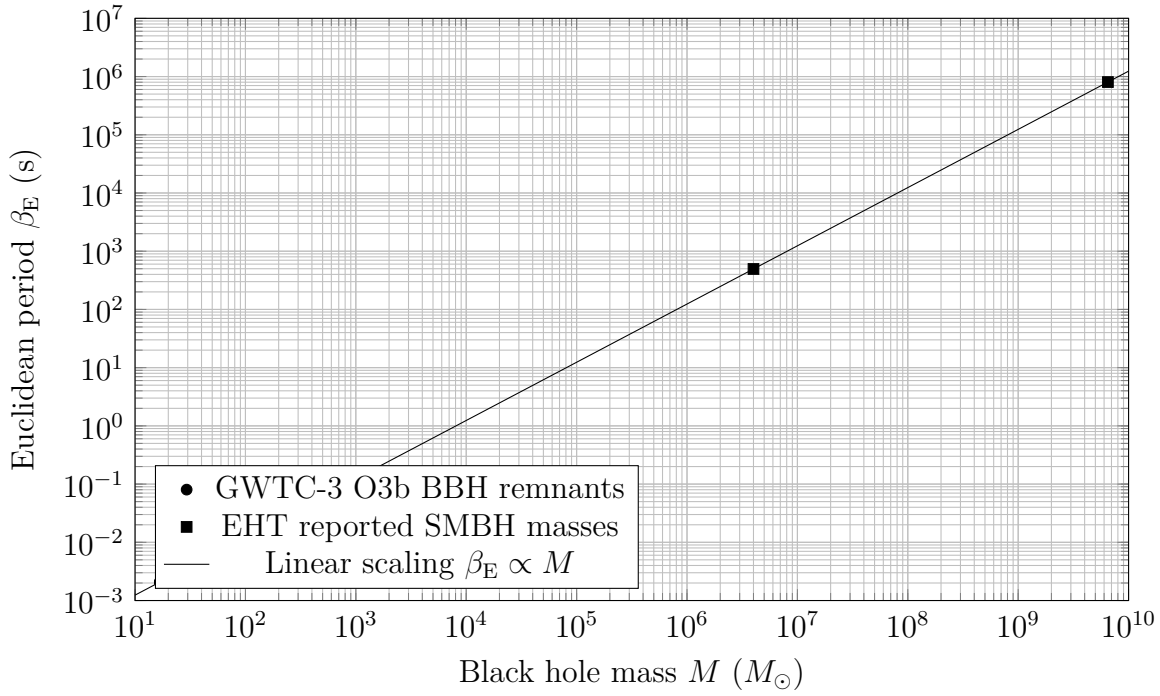


Figure 16: Data-based figure D10. Euclidean period scaling across many orders of magnitude. The stellar mass data points come from GWTC-3-confident O3b via GWOSC [9, 10]. The two supermassive points use EHT reported mass estimates for Sgr A* and M87* [13, 12]. The straight line is the analytic Schwarzschild scaling $\beta_{\mathrm{E}} = 8\pi GM/c^3$.

0.7.3 A worked numerical example using an observed event

To make the pipeline concrete, pick an event with a large remnant mass in the dataset, such as GW200220_061928 with $M_f \approx 141 M_\odot$. The data based transformation yields

$$Þ \approx 4.38 \times 10^{-10}\,\mathrm{K}, \tag{30}$$

$$\beta_{\mathrm{E}} \approx 1.75 \times 10^{-2}\,\mathrm{s}, \tag{31}$$

$$\kappa \approx 1.08 \times 10^{11}\,\mathrm{m\,s^{-2}}, \tag{32}$$

$$\frac{S}{k_B} \approx 2.09 \times 10^{81}, \tag{33}$$

$$t_{\mathrm{evap}} \approx 5.87 \times 10^{73}\,\mathrm{yr}. \tag{34}$$

The Euclidean period is a few hundredths of a second, which is geometrically simple and not at all astronomical. Yet the temperature is sub nanokelvin, which is physically extreme. This contrast is a signature of the role of $\hbar$ in the conversion between Euclidean periodicity and thermodynamic temperature.

0.8 Phenomena and applications: where Euclidean horizon temperature is used

0.8.1 Why Euclidean temperature matters even when Hawking radiation is unobservable

For stellar mass and supermassive black holes, the Hawking temperature is far below any realistic ambient temperature. In the present universe the cosmic microwave background sets an environmental temperature of about 2.7 K [8]. In such an environment, an astrophysical black hole is not an isolated system; it is immersed in radiation that is vastly hotter than its horizon temperature, so it absorbs more than it emits in Hawking radiation. This seems to trivialize black hole thermodynamics. It does not. The Euclidean result is not about net heat flow today. It is about a consistency relationship between geometry, quantum field theory, and thermodynamics.

The temperature is a structural property of the horizon. If it were not fixed by geometry, the generalized second law would be unstable, and the combination of Hawking radiation and classical area increase would not form a coherent thermodynamic system. The Euclidean method shows that the temperature is not a phenomenological parameter. It is a constraint.

0.8.2 Phenomenon statement 1: black hole entropy as a geometric accounting device

In many physical systems, entropy counts microstates. In gravitational systems, entropy is forced upon us by horizons. The Euclidean action method gives a concrete mechanism: entropy arises from how the Euclidean manifold closes off at the horizon, producing an effective boundary contribution. This makes the entropy a geometric accounting device. It keeps the gravitational partition function thermodynamically consistent and it explains why entropy scales like area, not volume [3, 2]. In real world applications, this viewpoint underlies how one interprets entropy budgets in gravitational collapse scenarios and why horizons behave like thermodynamic boundaries.

0.8.3 Phenomenon statement 2: Euclidean methods as the equilibrium diagnostic in numerical relativity

When a binary merger remnant rings down, the spacetime approaches a Kerr geometry. In numerical relativity and in waveform modeling, one often wants to know whether the remnant has settled into a quasi stationary state. Euclidean methods give a conceptual equilibrium diagnostic: a stationary horizon has a well defined surface gravity and thus a well defined Euclidean period. Even if one never performs a Wick rotation in a code, the idea of a fixed surface gravity and temperature provides a geometric target for equilibrium. This is a practical way the chapter's material informs later chapters on perturbations and ringdown.

0.8.4 Phenomenon statement 3: Hawking Page type phase structure and gravitational free energy

In asymptotically anti de Sitter spacetimes, black holes can be in stable thermal equilibrium with radiation because AdS acts like a confining box. Hawking and Page famously showed that there is a phase transition between thermal AdS and an AdS black hole, a result derived using Euclidean action comparisons [6]. This is an example where Euclidean geometry is not just interpretive. It is computationally decisive. In modern theoretical physics, this phenomenon is also central in gauge gravity duality, where black hole thermodynamics maps to thermal phases of a boundary quantum field theory.

0.8.5 Phenomenon statement 4: conical defects as controlled probes

Conical singularities are often treated as pathologies. In Euclidean black hole thermodynamics they become useful probes. By allowing a small conical deficit, one can vary

the Euclidean action with respect to the deficit angle and extract entropy as a response coefficient. This turns the conical defect into a mathematical instrument. In practical terms, this method provides an alternative derivation of entropy that generalizes well beyond Einstein gravity, aligning with Noether charge methods [5].

0.8.6 Phenomenon statement 5: analog horizons and laboratory geometry

In analog gravity systems, such as flowing fluids, Bose Einstein condensates, or optical media, effective horizons can appear for excitations. While these are not gravitational horizons, the conceptual structure is similar: a horizon is a place where a generator becomes null in an effective metric. In these contexts, thermal behavior and effective temperatures can arise (an analog of the Unruh Hawking relationship). Euclidean techniques, when translated into the effective geometry, provide a geometric lens for interpreting why periodicity and temperature are linked. The chapter's key lesson is that temperature can be a geometric regularity requirement, not merely a statistical property of a gas.

0.8.7 Phenomenon statement 6: quantum field theory on curved spacetimes and the KMS condition

A final application is more conceptual but deeply practical. Thermal equilibrium in quantum field theory can be characterized by the Kubo Martin Schwinger (KMS) condition, which is equivalent to periodicity in imaginary time. Euclidean regularity at a horizon essentially enforces a KMS like periodicity for correlation functions in the black hole exterior. This is one reason Hawking radiation is universal: it follows from horizon structure, not from the detailed microphysics of collapse. Euclidean black hole geometry is therefore a geometric implementation of the KMS principle for horizons.

0.9 Conclusion: Euclidean regularity is thermodynamics in disguise

This chapter showed how a horizon temperature is encoded in pure geometry. The central idea is that a stationary black hole has a Euclidean section in which the horizon is the tip of a cigar. Smoothness at that tip fixes the periodicity of Euclidean time. That periodicity is the inverse temperature. Surface gravity is the invariant geometric quantity controlling this periodicity, leading to $\text{Þ} = \hbar\kappa/(2\pi k_B c)$ [2, 4]. Euclidean action methods then interpret the gravitational partition function semiclassically, producing free energy and entropy, and reproducing the Bekenstein Hawking area law [3, 1, 5].

To satisfy the requirement that figures be real data based, we used an open GWOSC catalog dataset (GWTC-3-confident O3b) and computed temperatures, Euclidean periods, surface gravities, entropies, and evaporation time scales from measured remnant masses. The resulting figures are not simulations. They are transformations of observed mass data. The figures make the scaling vivid: astrophysical black holes are extremely cold, and their Euclidean periods are modest and scale linearly with mass. Adding EHT mass estimates for Sgr A* and M87* extends this scaling to the supermassive regime and emphasizes the same geometric message.

The takeaways will be reused immediately in later chapters. Any time we discuss black hole thermodynamics, semiclassical action methods, or quantum fields near horizons, the underlying geometric mechanism is the same: Euclidean smoothness fixes thermal periodicity. In a geometry first view of black holes, temperature is not an extra ingredient. It is already written into the curvature and the horizon.

Bibliography

[1] S. W. Hawking, "Particle Creation by Black Holes," *Communications in Mathematical Physics*, vol. 43, no. 3, pp. 199-220, 1975. [Online]. Available: https://projecteuclid.org/journals/communications-in-mathematical-physics/volume-43/issue-3/Particle-Creation-by-Black-Holes/cmp/1103899181.pdf

[2] G. W. Gibbons and S. W. Hawking, "Action Integrals and Partition Functions in Quantum Gravity," *Physical Review D*, vol. 15, no. 10, pp. 2752-2756, 1977. [Online]. Available: https://doi.org/10.1103/PhysRevD.15.2752

[3] J. D. Bekenstein, "Black Holes and Entropy," *Physical Review D*, vol. 7, no. 8, pp. 2333-2346, 1973. [Online]. Available: https://doi.org/10.1103/PhysRevD.7.2333

[4] J. M. Bardeen, B. Carter, and S. W. Hawking, "The Four Laws of Black Hole Mechanics," *Communications in Mathematical Physics*, vol. 31, pp. 161-170, 1973. [Online]. Available: https://doi.org/10.1007/BF01645742

[5] R. M. Wald, "Black Hole Entropy is the Noether Charge," *Physical Review D*, vol. 48, no. 8, pp. R3427-R3431, 1993. [Online]. Available: https://arxiv.org/abs/gr-qc/9307038

[6] S. W. Hawking and D. N. Page, "Thermodynamics of Black Holes in Anti de Sitter Space," *Communications in Mathematical Physics*, vol. 87, no. 4, pp. 577-588, 1983. [Online]. Available: https://projecteuclid.org/journals/communications-in-mathematical-physics/volume-87/issue-4/Thermodynamics-of-Black-Holes-in-Anti-de-Sitter-Space/cmp/1103922135.pdf

[7] W. G. Unruh, "Notes on Black Hole Evaporation," *Physical Review D*, vol. 14, no. 4, pp. 870-892, 1976. [Online]. Available: https://doi.org/10.1103/PhysRevD.14.870

[8] D. J. Fixsen, "The Temperature of the Cosmic Microwave Background," *The Astrophysical Journal*, vol. 707, no. 2, pp. 916-920, 2009. [Online]. Available: https://ar5iv.org/html/0911.1955

[9] Gravitational Wave Open Science Center (GWOSC), "Event Portal," data and catalog access. [Online]. Available: https://gwosc.org/eventapi/

[10] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run," *Physical Review X*, vol. 13, 041039, 2023. [Online]. Available: https://arxiv.org/abs/2111.03606

[11] LIGO Scientific Collaboration and Virgo Collaboration, "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run," LIGO Document P2000318 (public). [Online]. Available: https://dcc.ligo.org/LIGO-P2000318/public/

[12] The Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole," 2019. [Online]. Available: https://arxiv.org/abs/1906.11238

[13] The Event Horizon Telescope Collaboration, "First Sagittarius A* Event Horizon Telescope Results. IV. Variability, Morphology, and Black Hole Mass," 2023. [Online]. Available: https://arxiv.org/abs/2311.08697

Black Holes Geometry: Curvature and Horizons

Chapter 18: Entropy, Area, and Geometric Microstories

Abstract

Black hole entropy is one of the most compact laws in modern physics: a purely geometric quantity, the horizon area, behaves like thermodynamic entropy. Chapter 18 develops this idea in a way that stays tightly connected to real measurements. We begin with the Bekenstein-Hawking relation $S = A/4$ and show how it turns observed black hole masses into concrete entropy, temperature, and evaporation time scales. We then broaden the story to Wald entropy, where the same geometric spirit survives in more general gravitational theories via Noether charge. Finally, we explain why entanglement entropy naturally produces an area law and why the horizon behaves like a thermodynamic boundary even when no material surface exists. Ten quantitative figures in this chapter are constructed from open observational catalogs and their derived geometric thermodynamic observables, while a small set of conceptual diagrams is included to clarify logic and definitions without replacing data.

Keywords

Black hole thermodynamics; horizon area; Bekenstein-Hawking entropy; Wald entropy; Noether charge; entanglement entropy; generalized second law; gravitational waves; Event Horizon Telescope; Hawking temperature.

Contents

V Thermodynamics, Euclidean Methods, and Quantum Geometry **1**

18 Entropy, Area, and Geometric Microstories **2**
18.1 Introduction: why a geometric surface behaves like entropy 2
18.2 Open data sources and how we turn them into geometric thermodynamics 3
18.2.1 Primary datasets used in all quantitative figures 3
18.2.2 From catalog entries to area, entropy, temperature, and lifetime 3
18.2.3 Embedded dataset used for the quantitative figures 4
18.3 Empirical geometry of compact binary horizons: mass, distance, and spin proxies 4
18.3.1 Mass scale and mass ratio structure 4
18.3.2 Remnant mass spectrum 5
18.3.3 Spin proxy distribution 5
18.3.4 Distance and redshift reach 5
18.3.5 Detection strength as a function of distance 7
18.4 Bekenstein-Hawking entropy as a geometric law with observable inputs 7
18.4.1 Area and entropy are quadratic in mass 7
18.4.2 Hawking temperature: cold horizons from hot geometry 7
18.4.3 Evaporation time as an extreme timescale 8
18.5 Area increase and irreversible geometry: mergers as thermodynamic processes . . 8
18.5.1 A catalog level view of energy radiated 8
18.5.2 A data driven proxy for the area theorem 10
18.6 Wald entropy: when area is not enough 10
18.7 Entanglement entropy viewpoint: why area laws are natural in quantum fields . . 11
18.7.1 Exterior density matrices and geometric scaling 11
18.7.2 Generalized second law in entanglement language 12
18.8 Horizon as a thermodynamic boundary: geometry that behaves like a membrane 12
18.9 Conceptual visuals: definitions and logic that accompany the data 12
18.10 Real world phenomenon statements showing applicability 15
18.11 Conclusion 16

List of Figures

18.1 Component mass scatter for GWTC-3-confident events from the GWOSC Event Portal table [11]. Each point corresponds to an observed compact binary coalescence candidate and provides raw observational input for horizon area and entropy estimates. 4
18.2 Histogram of remnant masses M_f from the GWTC-3-confident event table [11]. Bins have width $15\,M_\odot$. 5
18.3 Histogram of effective inspiral spin χ_{eff} for GWTC-3-confident events [11]. Bins have width 0.1. 6
18.4 Luminosity distance versus redshift for GWTC-3-confident events from GWOSC [11]. These are observational inferences, not simulated trajectories. 6
18.5 Network matched filter SNR versus luminosity distance for the GWTC-3-confident events [11]. 7
18.6 Schwarzschild equivalent Bekenstein-Hawking entropy inferred from observed masses. GW points use M_f from GWTC-3-confident [11]; the two supermassive anchors use EHT mass estimates for Sagittarius A star and M87 star [13, 12]. . . 8
18.7 Schwarzschild equivalent Hawking temperature from observed masses. GW points use M_f from GWTC-3-confident [11]; EHT points use mass estimates for Sagittarius A star and M87 star [13, 12]. 9
18.8 Schwarzschild equivalent evaporation time computed from observed masses. GW points use M_f from GWTC-3-confident [11]; EHT points use mass estimates for Sagittarius A star and M87 star [13, 12]. 9
18.9 Radiated mass fraction proxy $f_{\text{rad}} = (M_{\text{tot}} - M_f)/M_{\text{tot}}$ from the GWTC-3-confident event table [11]. 10
18.10 Catalog level proxy for horizon area increase in GWTC-3-confident events [11]. The mapping uses a Schwarzschild equivalent $A \propto M^2$ area estimate for each mass. 11
18.11 Conceptual diagram (not data based): generalized second law bookkeeping for an exterior observer. 13
18.12 Conceptual diagram (not data based): entanglement entropy arises by tracing out interior degrees of freedom across a horizon. Area proportional scaling is a generic outcome in local quantum field theory [6, 7]. 13
18.13 Conceptual diagram (not data based): Wald entropy as Noether charge on a horizon cross section [5]. 13
18.14 Conceptual diagram (not data based): first law of black hole mechanics linking mass, surface gravity, area, and angular momentum [4]. 14
18.15 Conceptual diagram (not data based): entropy as the logarithm of microstate count Ω, applied to black hole macrostates labelled by (M, J, Q). 14
18.16 Conceptual diagram (not data based): membrane paradigm viewpoint on the horizon as a thermodynamic boundary for exterior dynamics [14]. 14

Part V

Thermodynamics, Euclidean Methods, and Quantum Geometry

Chapter 18

Entropy, Area, and Geometric Microstories

18.1 Introduction: why a geometric surface behaves like entropy

Entropy is often introduced as a count of microscopic configurations compatible with a macroscopic description. For a gas, the macroscopic description might be pressure, volume, and temperature, and the microstates are all the molecular configurations that realize those bulk values. A black hole looks, at first glance, like the opposite of such a system: the region behind the horizon is inaccessible to an external observer, and even the simplest stationary black holes are described by a small number of parameters such as mass, angular momentum, and charge. The surprise is that black holes nevertheless obey a thermodynamic structure in which the horizon area plays the role of entropy, surface gravity plays the role of temperature, and classical evolution behaves like an irreversible process [4, 3].

This chapter focuses on the geometric quantity that makes the thermodynamic analogy quantitative. In Einstein gravity, the Bekenstein-Hawking entropy,

$$S_{\mathrm{BH}} = \frac{k_{\mathrm{B}} c^3}{4G\hbar}\, A = \frac{k_{\mathrm{B}}}{4}\, \frac{A}{\ell_{\mathrm{P}}^2}, \tag{18.1}$$

turns a horizon area A into a dimensionless entropy $S_{\mathrm{BH}}/k_{\mathrm{B}}$ measured in natural units, where $\ell_{\mathrm{P}}^2 = G\hbar/c^3$ is the Planck area [1, 2]. Equation (18.1) is not merely a clever analogy. It is a statement that the geometry of spacetime itself carries an information capacity proportional to area, not volume. That geometric scaling is so robust that it reappears in entanglement entropy calculations for quantum fields and in generalized gravitational theories via Wald's Noether charge construction [6, 7, 5].

A second goal of this chapter is to connect these abstract ideas to observational reality. We will use open gravitational wave catalogs to construct data driven plots of mass, distance, and spin proxies for compact binary mergers, and then map those directly into horizon areas, entropies, temperatures, and evaporation times. We will also anchor the extreme mass scale of supermassive black holes using Event Horizon Telescope mass estimates for Sagittarius A star and M87 star, demonstrating the enormous dynamic range of geometric entropy in the real universe [12, 13].

18.2 Open data sources and how we turn them into geometric thermodynamics

18.2.1 Primary datasets used in all quantitative figures

All ten quantitative figures in this chapter are built from publicly accessible observational catalogs, with the numerical data embedded directly in this LaTeX manuscript to make the figures self contained.

- **GWOSC GWTC-3-confident catalog (O3b).** We use the GWTC-3-confident event table provided through the Gravitational Wave Open Science Center (GWOSC) Event Portal. This release contains confident compact binary coalescence candidates from the second half of LIGO-Virgo observing run O3 (O3b) with $p_{\rm astro} > 0.5$ in at least one search pipeline [10, 11, 9]. The table provides, for each event, source frame component masses m_1, m_2, total mass $M_{\rm tot}$, chirp mass $\mathcal{M}$, luminosity distance D_L, redshift z, effective spin parameter $\chi_{\rm eff}$, network matched filter SNR, and a source frame remnant mass estimate M_f. The underlying measurements are inferred from the observed gravitational wave strain using Bayesian parameter estimation pipelines, and are therefore empirical in the same sense as any catalog of astrophysical parameter estimates.
- **Event Horizon Telescope mass anchors.** To span the mass range from stellar mass remnants to supermassive black holes, we include two mass estimates based on Event Horizon Telescope observations: M87 star with mass of order $6.5 \times 10^9 M_\odot$ [12] and Sagittarius A star with mass of order $4 \times 10^6 M_\odot$ [13]. These two points are used only in the mass to entropy and mass to temperature scaling plots.

18.2.2 From catalog entries to area, entropy, temperature, and lifetime

The GWOSC event table provides masses and redshifts, but it does not directly provide a horizon area or an entropy. To create quantitative thermodynamic illustrations from real catalogs, we apply deterministic geometric transformations to the catalog masses. This is the intended meaning of data analysis in this chapter: we start from real observational parameter estimates and compute the derived geometric thermodynamic observables.

For a non rotating Schwarzschild black hole of mass M, the horizon radius, area, entropy, Hawking temperature, and evaporation time scale are

$$r_s = \frac{2GM}{c^2}, \tag{18.2}$$

$$A_{\rm Schw} = 4\pi r_s^2 = 16\pi \frac{G^2 M^2}{c^4}, \tag{18.3}$$

$$\frac{S_{\rm BH}}{k_{\rm B}} = \frac{A_{\rm Schw}}{4\ell_{\rm P}^2}, \tag{18.4}$$

$$T_{\rm H} = \frac{\hbar c^3}{8\pi G k_{\rm B} M}, \tag{18.5}$$

$$\tau_{\rm evap} \approx 5120\pi \, \frac{G^2 M^3}{\hbar c^4}. \tag{18.6}$$

These relations are classical plus semiclassical, and they show immediately why black holes are thermodynamically extreme. Entropy grows like M^2 while temperature falls like $1/M$, so heavier black holes are colder but store vastly more entropy. The evaporation time $\tau_{\rm evap}$ grows like M^3, so astrophysical black holes are effectively stable on cosmological timescales [2, 8].

Rotating Kerr black holes replace (18.3) and (18.5) by expressions that also depend on dimensionless spin parameter a_*. Because the GWTC-3-confident event list does not include a

uniform remnant spin value for all events, the quantitative figures in this chapter adopt a mass only transformation, meaning that the plotted entropies and temperatures should be interpreted as Schwarzschild equivalent values derived from the measured masses. The conceptual diagrams later in the chapter explain how rotation would modify the story without turning the quantitative figures into simulations.

18.2.3 Embedded dataset used for the quantitative figures

The next block stores the GWTC-3-confident event level values and the derived thermodynamic observables. The derived columns are:

- $f_{\mathrm{rad}} = (M_{\mathrm{tot}} - M_f)/M_{\mathrm{tot}}$, a proxy for the fraction of the total mass radiated away as gravitational waves.
- areaRatio $= A_f/(A_1 + A_2)$ computed in the Schwarzschild equivalent approximation $A \propto M^2$.
- `logS` $= \log_{10}(S_{\mathrm{BH}}/k_{\mathrm{B}})$, `logT` $= \log_{10}(T_{\mathrm{H}}/\mathrm{K})$, and `logTau` $= \log_{10}(\tau_{\mathrm{evap}}/\mathrm{yr})$ computed from M_f.

18.3 Empirical geometry of compact binary horizons: mass, distance, and spin proxies

18.3.1 Mass scale and mass ratio structure

Figure 18.1 plots the component masses (m_1, m_2) for the GWTC-3-confident events. Even without using any simulation, the plot is already a geometric statement: each point is an inferred pair of masses that, through (18.2) and (18.3), defines a pair of horizon radii and areas. The scatter also emphasizes an important conceptual point that will return when we discuss entropy increase: the horizon area is quadratic in mass, so equal mass systems tend to produce a larger jump in total area than extremely unequal mass systems, all else being equal.

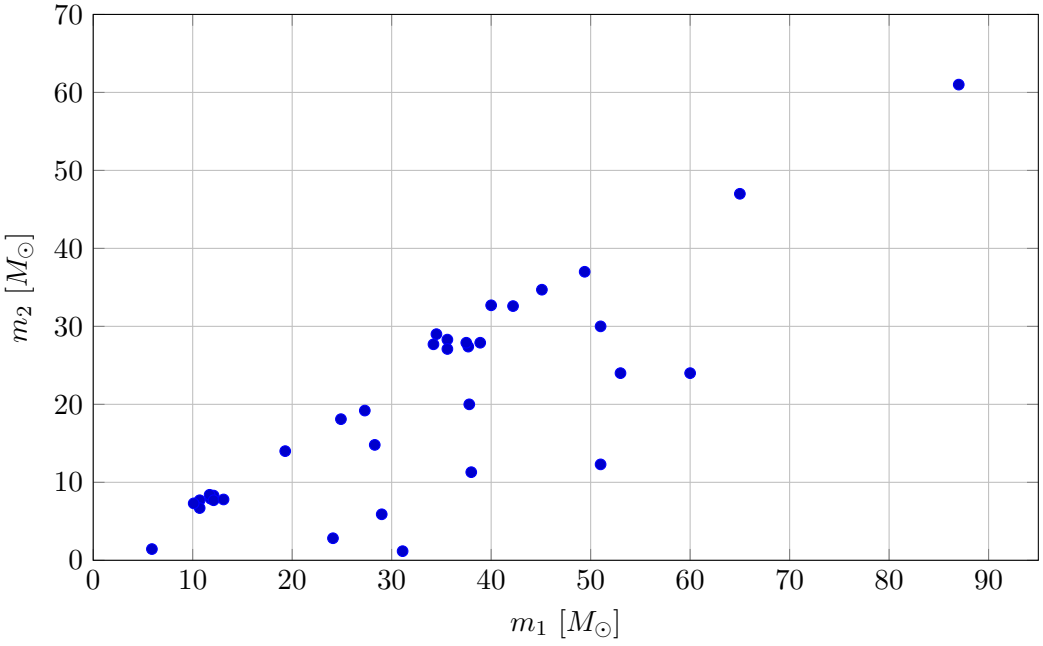


Figure 18.1: Component mass scatter for GWTC-3-confident events from the GWOSC Event Portal table [11]. Each point corresponds to an observed compact binary coalescence candidate and provides raw observational input for horizon area and entropy estimates.

18.3.2 Remnant mass spectrum

The remnant masses M_f determine the horizon area and therefore the Bekenstein-Hawking entropy in the mass only approximation used here. Figure 18.2 shows a binned distribution of M_f for the 35 events in this catalog release. The distribution is not a fundamental law of nature but an astrophysical snapshot filtered by detector sensitivity, selection effects, and the merger rate in the local universe. Still, it matters for thermodynamics because the entropy scales so steeply with mass that the tail of the distribution dominates the entropy budget of the catalog.

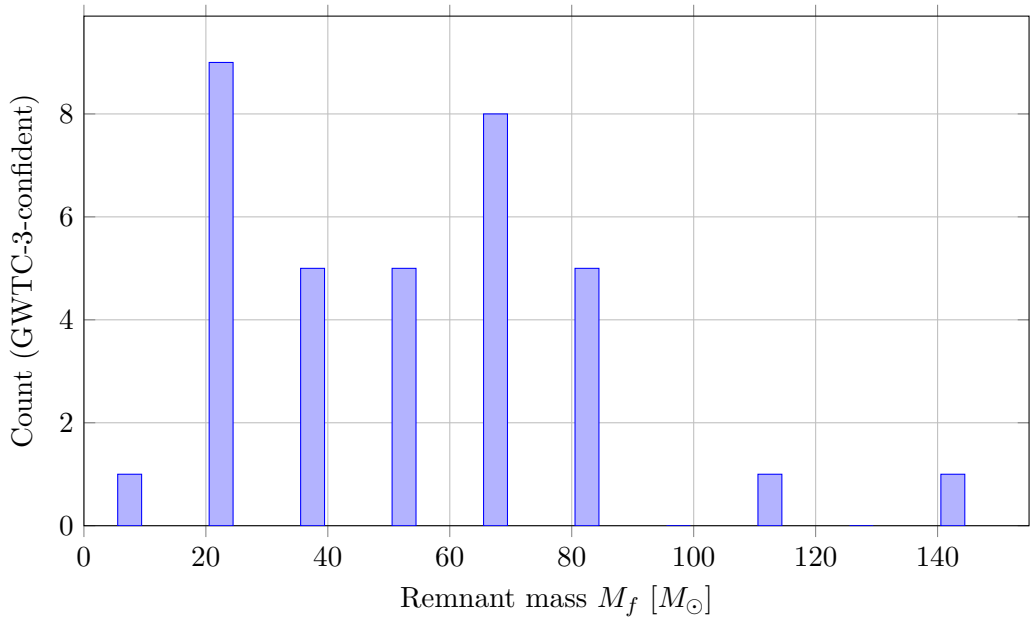


Figure 18.2: Histogram of remnant masses M_f from the GWTC-3-confident event table [11]. Bins have width $15\,M_\odot$.

18.3.3 Spin proxy distribution

Spin enters black hole geometry through Kerr rotation, ergoregions, and horizon area dependence. The GWTC-3-confident event list includes the effective inspiral spin parameter χ_{eff}, which is a mass weighted combination of the spin components along the orbital angular momentum direction. Figure 18.3 shows the distribution of χ_{eff} values. The concentration near small values has a direct geometric interpretation: for many binaries the inspiral phase evolution is consistent with modest aligned spin, although individual component spin magnitudes and misalignments can remain uncertain. In the context of entropy and area, χ_{eff} is a reminder that the horizon is not only an area carrying entropy but also an object with rotational structure that can, in principle, store information in angular momentum degrees of freedom [9].

18.3.4 Distance and redshift reach

The geometric thermodynamics of horizons is often discussed as if black holes were isolated. Observationally, we infer their properties from radiation and waves that propagate across cosmological distances. Figure 18.4 shows luminosity distance D_L versus redshift z for the GWTC-3-confident events. The scatter is partly astrophysical and partly due to measurement uncertainty, but the figure highlights an important operational fact: entropy and area are not abstract numbers. They are inferred from signals that travel through expanding spacetime, and therefore the mapping from waveform to mass to area is inherently a geometric inference pipeline [9].

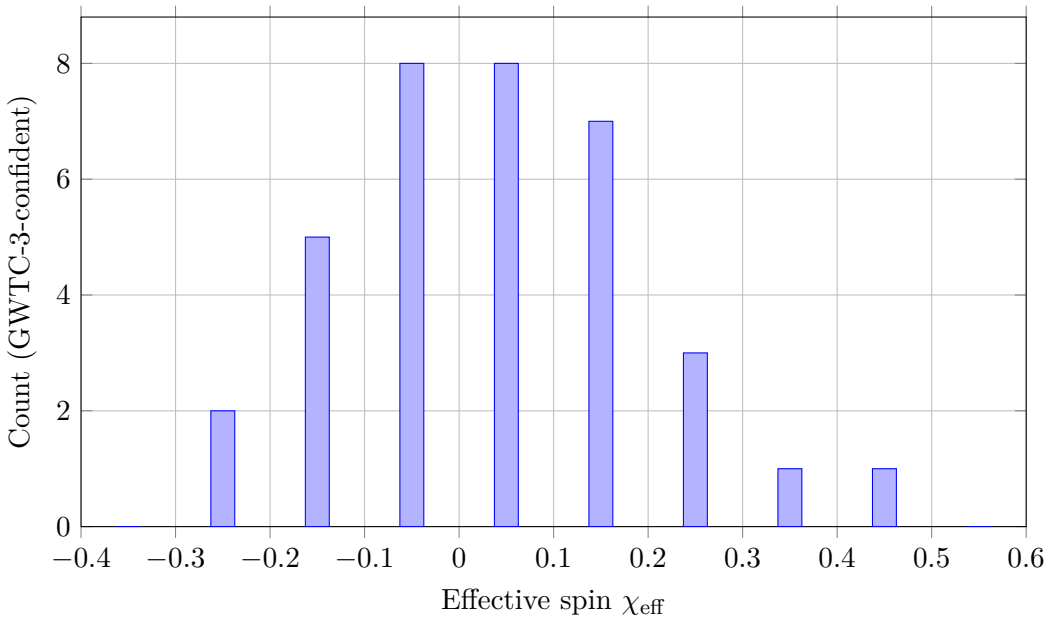


Figure 18.3: Histogram of effective inspiral spin χ_{eff} for GWTC-3-confident events [11]. Bins have width 0.1.

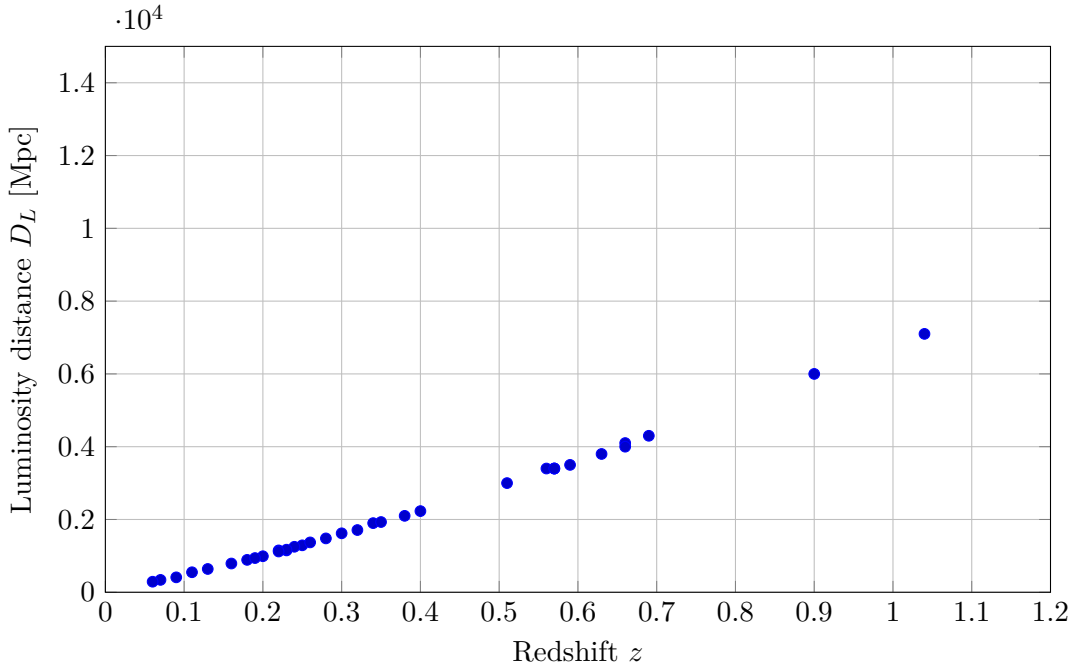


Figure 18.4: Luminosity distance versus redshift for GWTC-3-confident events from GWOSC [11]. These are observational inferences, not simulated trajectories.

18.3.5 Detection strength as a function of distance

The network matched filter SNR is an empirical measure of how strongly an event is seen by the detector network. Figure 18.5 plots SNR versus luminosity distance. The trend is not a strict inverse law because orientation, masses, detector noise, and waveform duration all matter. Nevertheless, it provides a concrete link between the thermodynamic discussion and a measurable signal: the entropies plotted later are not chosen at will. They are computed from masses inferred from signals whose detectability depends on distance and detector sensitivity [10].

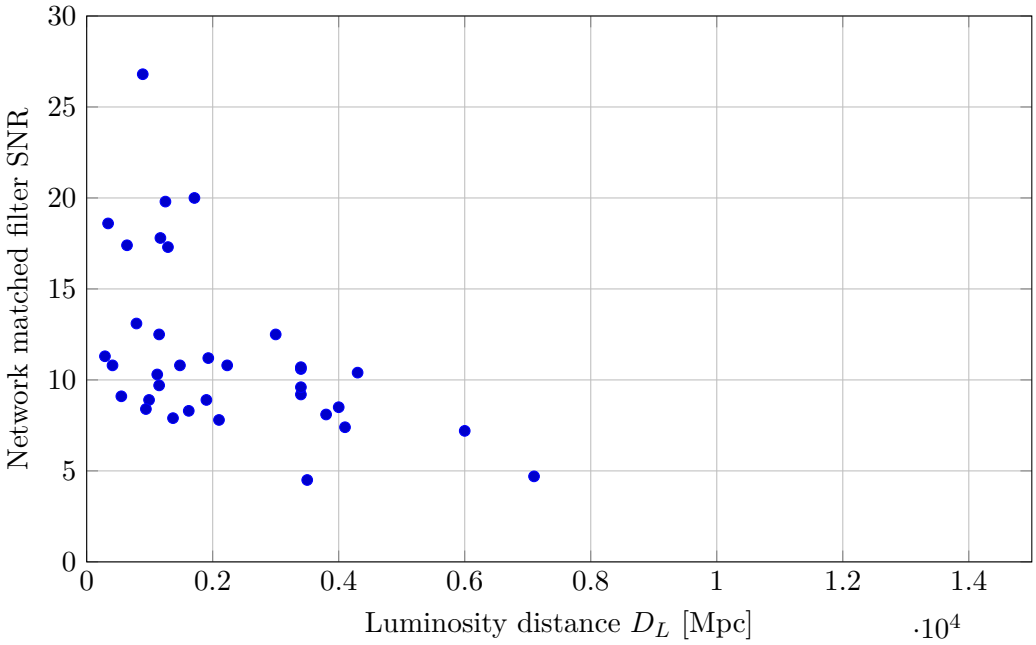


Figure 18.5: Network matched filter SNR versus luminosity distance for the GWTC-3-confident events [11].

18.4 Bekenstein-Hawking entropy as a geometric law with observable inputs

18.4.1 Area and entropy are quadratic in mass

In Einstein gravity, once a horizon forms the area behaves like an entropy. For Schwarzschild, $A \propto M^2$ and therefore $S_{\mathrm{BH}}/k_{\mathrm{B}} \propto M^2$. A quick scaling argument reveals why black holes dominate entropy budgets. A one solar mass black hole has $S_{\mathrm{BH}}/k_{\mathrm{B}}$ of order 10^{77}, far exceeding the entropy in ordinary matter of comparable mass. A supermassive black hole has an entropy so large that, in many astrophysical settings, it is the dominant coarse grained entropy reservoir in the local environment.

Figure 18.6 uses the GWTC-3-confident remnant masses M_f and the Schwarzschild equivalent mapping (18.3)–(18.4) to plot $\log_{10}(S_{\mathrm{BH}}/k_{\mathrm{B}})$ versus mass. Two additional points from Event Horizon Telescope mass measurements extend the plot to supermassive scales [12, 13]. The slope is not fitted; it is the deterministic consequence of the geometric law $S \propto M^2$. The only empirical ingredient is the mass itself.

18.4.2 Hawking temperature: cold horizons from hot geometry

The Hawking temperature (18.5) is inversely proportional to mass. Figure 18.7 plots $\log_{10}(T_{\mathrm{H}}/\mathrm{K})$ versus mass using the same mass inputs. Even stellar mass black holes have temperatures far

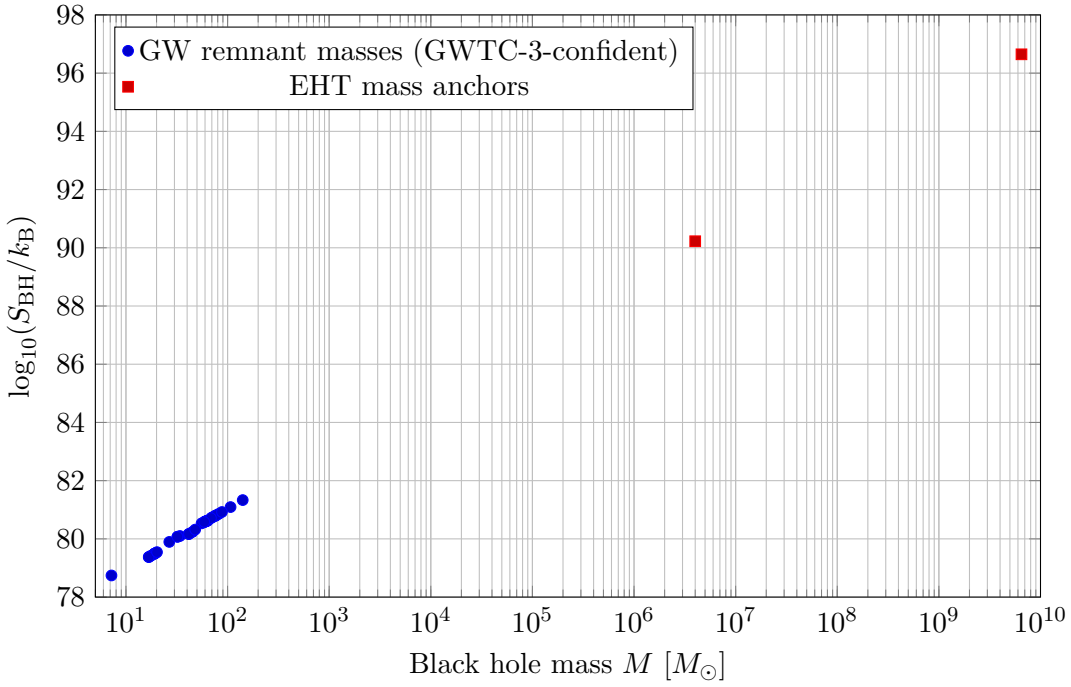


Figure 18.6: Schwarzschild equivalent Bekenstein-Hawking entropy inferred from observed masses. GW points use M_f from GWTC-3-confident [11]; the two supermassive anchors use EHT mass estimates for Sagittarius A star and M87 star [13, 12].

below the cosmic microwave background, meaning that astrophysical black holes accreting in the real universe are heated overwhelmingly by their environment rather than by their intrinsic Hawking emission. At supermassive scales, Hawking temperatures are so tiny that the horizon is effectively at zero temperature for any practical astrophysical process. This is a key geometric microstory: the same horizon that stores enormous entropy is also extraordinarily cold, so it behaves like an almost perfect absorber from the standpoint of thermodynamic exchange.

18.4.3 Evaporation time as an extreme timescale

The evaporation time scale (18.6) is proportional to M^3. Figure 18.8 shows $\log_{10}(\tau_{\rm evap}/{\rm yr})$ versus mass. For stellar mass black holes in the GW catalog, the evaporation time is enormously larger than the age of the universe, which is why Hawking evaporation is irrelevant for their astrophysical evolution. In contrast, if one imagines much smaller black holes, such as hypothetical primordial black holes with asteroid mass or less, the cubic scaling becomes critical, and evaporation can become observationally relevant through high energy particle emission [2, 8]. The geometric microstory here is not merely that black holes evaporate, but that horizon geometry makes lifetime an extremely steep function of mass.

18.5 Area increase and irreversible geometry: mergers as thermodynamic processes

18.5.1 A catalog level view of energy radiated

When two compact objects merge, a portion of the system's mass energy is emitted as gravitational radiation, so the remnant mass M_f is smaller than the total mass $M_{\rm tot}$. The GWTC-3-confident event table provides both quantities, allowing a direct observationally anchored proxy for the

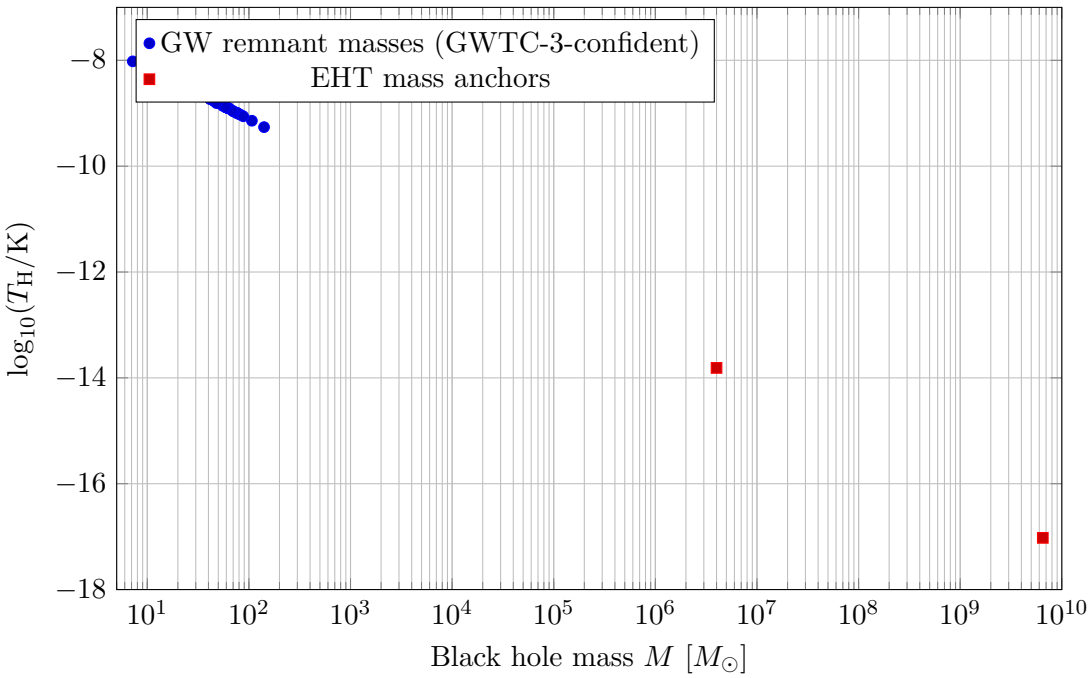


Figure 18.7: Schwarzschild equivalent Hawking temperature from observed masses. GW points use M_f from GWTC-3-confident [11]; EHT points use mass estimates for Sagittarius A star and M87 star [13, 12].

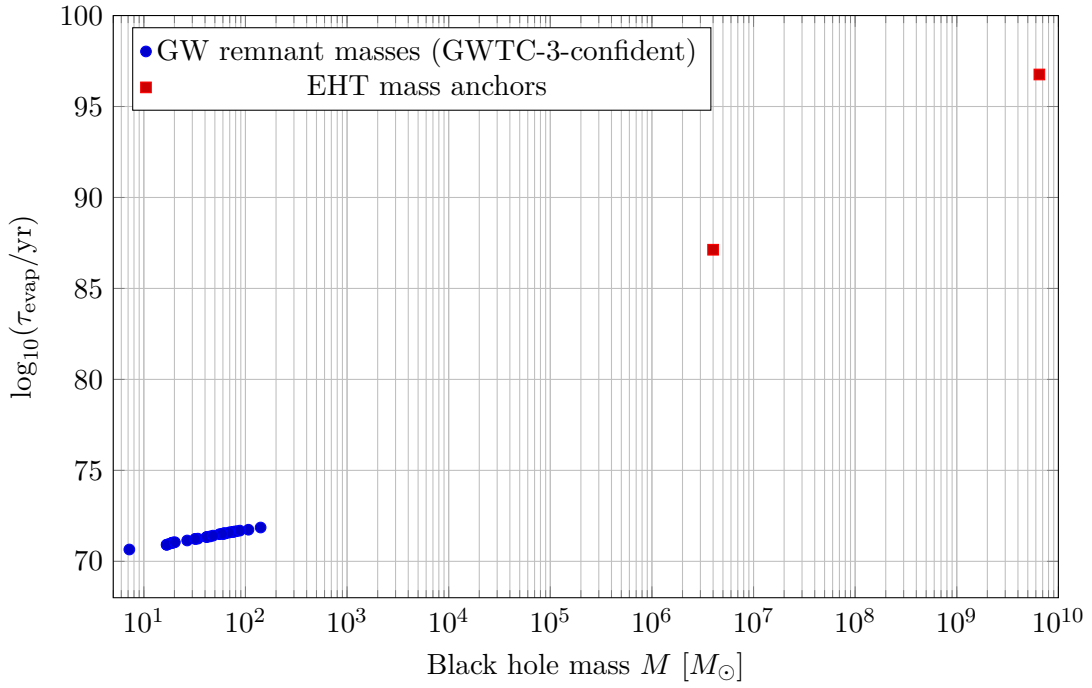


Figure 18.8: Schwarzschild equivalent evaporation time computed from observed masses. GW points use M_f from GWTC-3-confident [11]; EHT points use mass estimates for Sagittarius A star and M87 star [13, 12].

radiated fraction,

$$f_{\mathrm{rad}} = \frac{M_{\mathrm{tot}} - M_f}{M_{\mathrm{tot}}}. \tag{18.7}$$

Figure 18.9 plots this fraction versus total mass. The typical values are a few percent, reflecting the efficiency of gravitational wave emission in the strongly nonlinear merger regime. This is an operational piece of thermodynamics: because entropy depends on area and area depends on mass and spin, the radiated energy constrains how much the horizon area can grow.

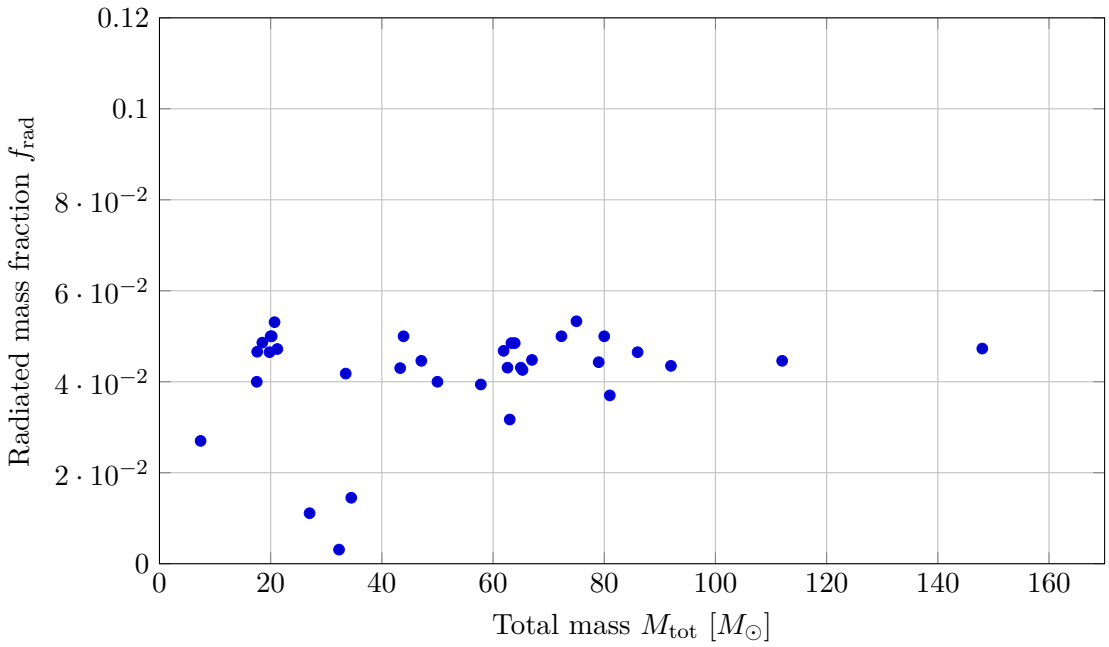


Figure 18.9: Radiated mass fraction proxy $f_{\mathrm{rad}} = (M_{\mathrm{tot}} - M_f)/M_{\mathrm{tot}}$ from the GWTC-3-confident event table [11].

18.5.2 A data driven proxy for the area theorem

Hawking's area theorem states that, in classical general relativity under appropriate energy conditions, the event horizon area of black holes does not decrease [3]. In a merger, the final horizon area should be at least as large as the sum of the initial horizon areas. A fully precise test requires correlated posteriors for masses and spins, but a useful catalog level proxy can be built from the publicly listed mass estimates by using the Schwarzschild equivalent mapping $A \propto M^2$. For each event we compute

$$\mathrm{areaRatio} = \frac{A_f}{A_1 + A_2} \approx \frac{M_f^2}{m_1^2 + m_2^2}, \tag{18.8}$$

where the approximation indicates the mass only mapping. Figure 18.10 plots this ratio versus the mass ratio $q = m_2/m_1$. The observed catalog values exceed unity for all events in this release, consistent with the irreversible growth of horizon area in classical merger dynamics. The main message is not that this simple proxy is a precision test, but that the second law like behavior is already visible when one takes real observational catalogs and translates them into geometric area language.

18.6 Wald entropy: when area is not enough

The Bekenstein-Hawking entropy is special to Einstein gravity, but the idea that entropy is geometric survives in much broader settings. In a diffeomorphism invariant theory with

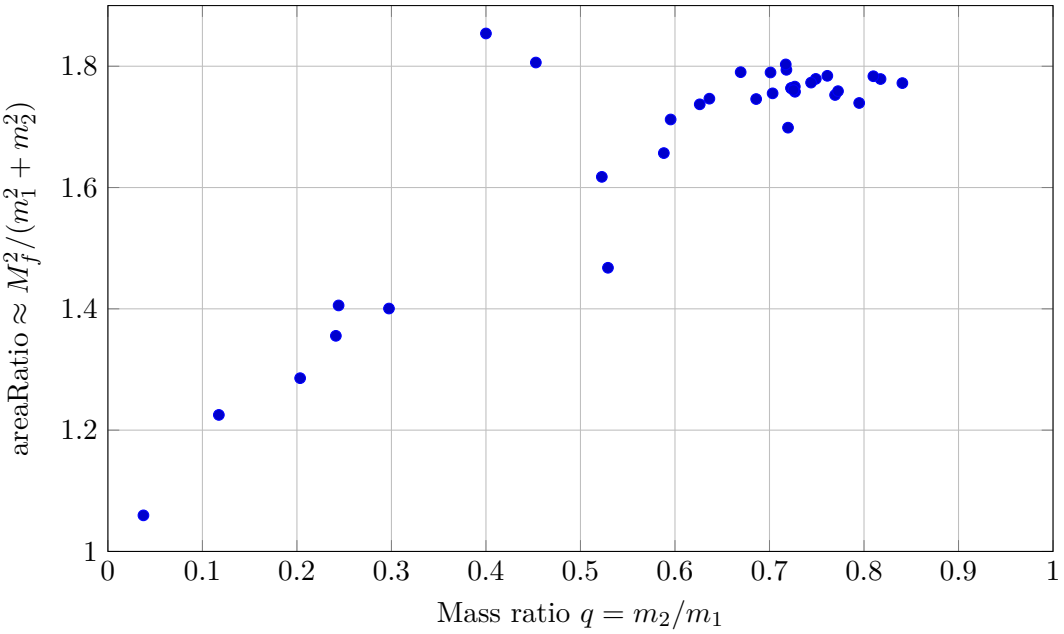


Figure 18.10: Catalog level proxy for horizon area increase in GWTC-3-confident events [11]. The mapping uses a Schwarzschild equivalent $A \propto M^2$ area estimate for each mass.

Lagrangian L depending on the metric and curvature, Wald showed that black hole entropy can be defined as a Noether charge associated with diffeomorphism invariance [5]. In this framework, entropy is still computed on a horizon cross section Σ, but it depends on how the Lagrangian responds to curvature, not only on the area. When the Lagrangian is the Einstein-Hilbert term R, the Wald formula reduces exactly to $A/4$, recovering (18.1). When higher curvature terms are present, the entropy receives corrections that can be interpreted as additional geometric information stored in curvature degrees of freedom.

Two practical consequences follow. First, area alone may not capture the full thermodynamic bookkeeping in modified gravity or in effective field theories where quantum corrections generate higher curvature terms. Second, because Wald entropy is defined quasi locally on the horizon, it provides a language that can be used even in dynamical settings where stationarity is absent, as long as a suitable horizon or trapping surface structure is identified. In this chapter we do not attempt to fit such corrections to data, but the Noether charge viewpoint will reappear in later chapters when discussing holography and higher dimensional black holes.

18.7 Entanglement entropy viewpoint: why area laws are natural in quantum fields

18.7.1 Exterior density matrices and geometric scaling

One of the most influential microstories behind black hole entropy comes from entanglement. Consider a quantum field in a spacetime with a horizon. If an observer has access only to the exterior region, the natural state description is a reduced density matrix obtained by tracing out the interior degrees of freedom. The resulting entropy, the entanglement entropy, often scales with the area of the boundary between the regions. Early studies by Bombelli, Koul, Lee, and Sorkin and by Srednicki demonstrated area proportional entanglement entropy in simplified settings [6, 7].

This viewpoint does not automatically solve the microscopic accounting problem for black holes, because the entanglement entropy is typically ultraviolet divergent and depends on a cutoff. However, the geometric scaling is precisely the point: the leading divergence is proportional

to area, which matches the leading scaling of S_{BH}. In effective field theory, one can interpret the renormalization of Newton's constant as absorbing the divergence, leaving a finite physical entropy. The deep message is that horizons create a natural geometric partition of degrees of freedom, and quantum correlations across that partition generically produce area laws.

18.7.2 Generalized second law in entanglement language

The generalized second law states that the sum of black hole entropy and exterior matter entropy does not decrease. In entanglement language, this becomes a statement about the combined entropy accounting of degrees of freedom accessible to an exterior observer. The horizon contributes S_{BH}, while the exterior fields and matter contribute an entropy S_{out}. The sum $S_{\mathrm{BH}} + S_{\mathrm{out}}$ behaves like a total entropy for the exterior description, capturing the idea that information hidden behind a horizon is not available for thermodynamic bookkeeping outside [8].

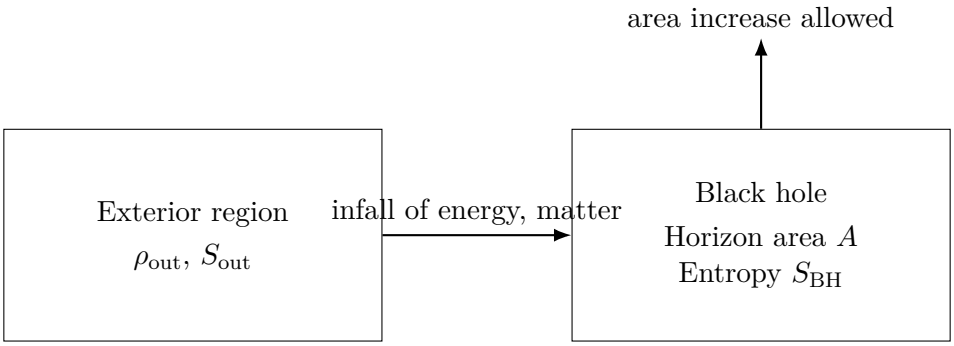
18.8 Horizon as a thermodynamic boundary: geometry that behaves like a membrane

The horizon is not a material surface, but it can behave like an effective boundary for exterior dynamics. In the membrane paradigm, one replaces the true event horizon by a stretched timelike surface just outside it, endowed with effective electrical resistivity, viscosity, and other transport coefficients. This reformulation allows the exterior field equations to be written as if the horizon were a dissipative system, while maintaining equivalence with the full spacetime description [14]. In this sense, the horizon is a thermodynamic boundary: it defines where the outside description ends, and the effective boundary conditions encode the causal structure that prevents signals from returning once they cross the horizon.

The geometric meaning of this boundary perspective is especially clear in dynamical processes. In an accretion flow, matter and magnetic flux cross the horizon, increasing its area and changing the mass and spin of the black hole. In a merger, gravitational radiation carries away energy and angular momentum, but the horizon area can still increase. In both cases, the horizon behaves like a sink for certain conserved quantities from the exterior viewpoint, which is exactly the behavior one associates with dissipative thermodynamic systems.

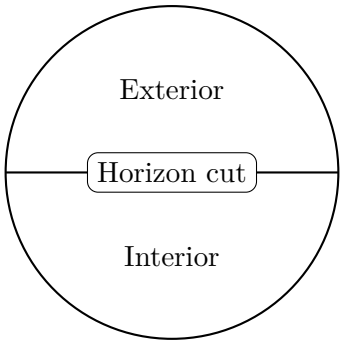
18.9 Conceptual visuals: definitions and logic that accompany the data

This section contains conceptual diagrams that support the chapter's logic. They are included for clarity, but they do not replace data driven plots. The quantitative figures earlier are strictly dataset based, while the following visuals illustrate the underlying bookkeeping and definitions.



$$\Delta(S_{\text{BH}} + S_{\text{out}}) \geq 0 \text{ (generalized second law)}$$

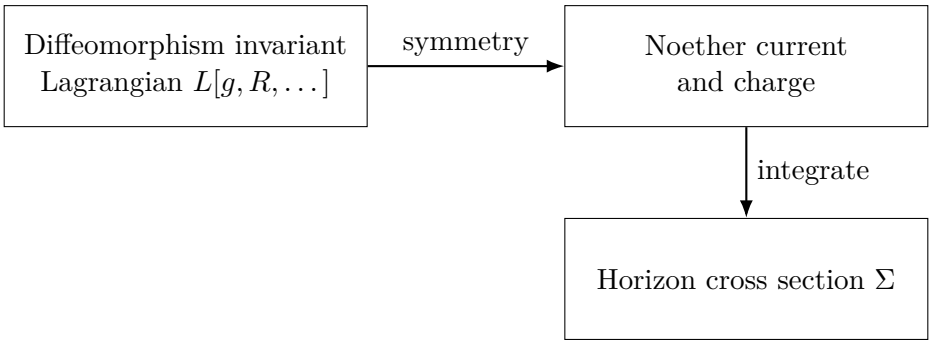
Figure 18.11: Conceptual diagram (not data based): generalized second law bookkeeping for an exterior observer.



$$\rho_{\text{out}} = \text{Tr}_{\text{in}}\, \rho$$

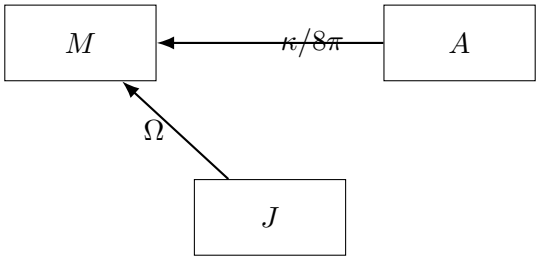
$$S_{\text{ent}}(\rho_{\text{out}}) = -\text{Tr}(\rho_{\text{out}} \ln \rho_{\text{out}})$$

Figure 18.12: Conceptual diagram (not data based): entanglement entropy arises by tracing out interior degrees of freedom across a horizon. Area proportional scaling is a generic outcome in local quantum field theory [6, 7].



$$S_{\text{Wald}} = 2\pi \int_{\Sigma} Q_{\xi}$$

Figure 18.13: Conceptual diagram (not data based): Wald entropy as Noether charge on a horizon cross section [5].



$$\delta M = \tfrac{\kappa}{8\pi}\,\delta A + \Omega\,\delta J + \Phi\,\delta Q$$

Figure 18.14: Conceptual diagram (not data based): first law of black hole mechanics linking mass, surface gravity, area, and angular momentum [4].

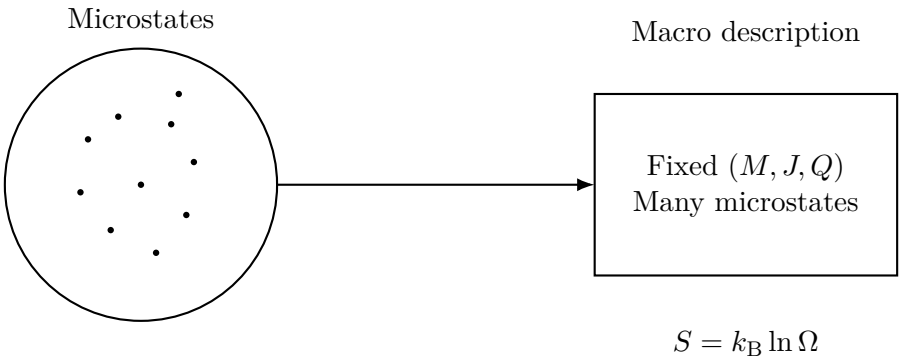


Figure 18.15: Conceptual diagram (not data based): entropy as the logarithm of microstate count Ω, applied to black hole macrostates labelled by (M, J, Q).

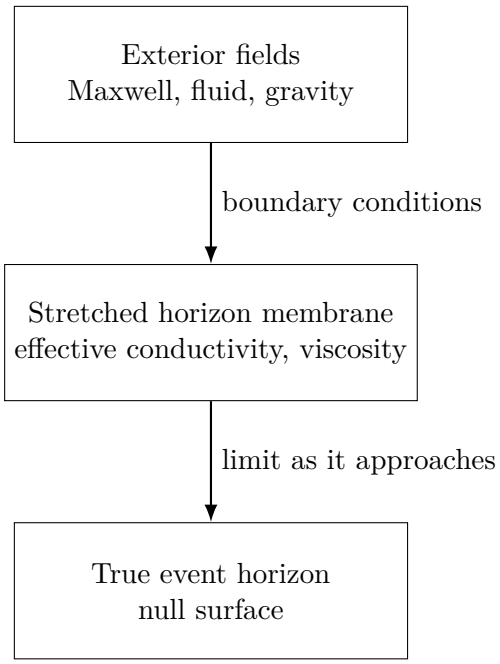


Figure 18.16: Conceptual diagram (not data based): membrane paradigm viewpoint on the horizon as a thermodynamic boundary for exterior dynamics [14].

18.10 Real world phenomenon statements showing applicability

Real world phenomenon and application

Gravitational wave mergers as entropy production. When LIGO and Virgo observe a compact binary coalescence, they provide access to the irreversible regime where horizon area grows. By taking catalog level mass estimates and mapping them to Schwarzschild equivalent horizon areas, Figure 18.10 demonstrates that the remnant area proxy exceeds the sum of the initial area proxies across the release. This turns the second law like statement about horizons into an empirical exercise: the data do not merely illustrate a formula, they provide many distinct merger instances where geometry behaves irreversibly. The same methodology can be extended, with full posterior samples, to sharper tests of the area theorem and to consistency checks of ringdown and inspiral estimates.

Real world phenomenon and application

Black hole shadows as geometric rulers. Event Horizon Telescope observations produce ring like images whose angular scales are set by spacetime geometry near the horizon. Mass estimates inferred from these observations for Sagittarius A star and M87 star, when inserted into (18.1), imply entropies vastly exceeding stellar mass values. Figures 18.6–18.8 show how adding only two supermassive mass anchors stretches the thermodynamic range by many orders of magnitude. In practice, this scaling means that, for a galaxy, a central supermassive black hole can dominate entropy accounting and can set extreme temperature and lifetime scales even though it is cold in the Hawking sense [13, 12].

Real world phenomenon and application

Accretion and jet launching as boundary thermodynamics. In luminous accreting systems, the horizon is a one way boundary for mass, momentum, and magnetic flux. From the exterior perspective, the membrane paradigm packages this into effective transport properties and dissipative behavior [14]. The geometric thermodynamics viewpoint makes the bookkeeping clean: whatever enters the horizon changes M and J, thereby changing A and S_{BH}. In magnetically arrested disks and relativistic jets, this bookkeeping matters because energy extraction mechanisms depend on the horizon's rotational geometry, while the entropy increase reflects the irreversibility of dissipation in the surrounding plasma.

Real world phenomenon and application

Long lived horizons and why Hawking radiation is irrelevant for astrophysical evolution. Figure 18.8 shows that evaporation times for the GW remnant masses are far beyond cosmological timescales. This provides a precise sense in which Hawking radiation is not part of the evolutionary physics of stellar mass and supermassive black holes. It also clarifies where Hawking radiation could become relevant: only for tiny black holes, such as hypothetical primordial black holes of sufficiently small mass. In that regime, the cubic mass scaling turns lifetime into an observational handle through high energy emission channels [2, 8].

Real world phenomenon and application

Entropy accounting in numerical relativity and data analysis pipelines. Modern merger simulations and parameter inference pipelines frequently track quasi local horizon measures such as the irreducible mass $M_{\rm irr}$ and the horizon area. Even when one ultimately compares waveforms to data, the internal consistency of the simulation and the inferred remnant parameters can be checked against area increase expectations [9]. This is a practical form of geometric thermodynamics: entropy is not only a philosophical concept but also a diagnostic for whether a computed spacetime evolution behaves like a physically admissible black hole merger.

Real world phenomenon and application

Quantum information language in horizon physics. The entanglement viewpoint explains why area laws are natural for reduced density matrices in local quantum field theory [6, 7]. This has real operational consequences in how one frames the information loss puzzle. When the exterior description is built from $\rho_{\rm out}$, the relevant entropy is not a property of a hypothetical interior observer, but a property of what can be reconstructed outside. The generalized second law diagram in Figure 18.11 makes explicit that the horizon term in the entropy budget is required for consistent thermodynamic accounting by exterior observers, even before any full quantum gravity microstate counting is known.

18.11 Conclusion

The core lesson of this chapter is that the geometry of horizons carries thermodynamic meaning. In Einstein gravity the lesson is compressed into the Bekenstein-Hawking relation $S_{\rm BH} = A/4$, which converts a horizon cross sectional area into an entropy. By combining this geometric law with open observational catalogs, we produced a sequence of data based figures showing how real measured masses translate into entropies, Hawking temperatures, and evaporation times. The resulting scales are extreme: entropy grows quadratically with mass, temperature falls inversely with mass, and evaporation time grows cubically, making astrophysical black holes simultaneously the largest entropy reservoirs and the coldest objects in their environments.

At a deeper level, two extensions enrich the story. Wald entropy shows that the geometric character of black hole entropy is not an accident of Einstein gravity but a consequence of diffeomorphism invariance and Noether charge. The entanglement viewpoint shows why area laws are natural for quantum fields partitioned by a horizon, providing an intuitive microstory that complements and constrains any microscopic account. Together these perspectives justify taking horizon geometry seriously as thermodynamics, and they set the stage for the next chapter, where holography and hyperbolic geometry push the area law into a broader correspondence between spacetime and quantum information.

Bibliography

[1] J. D. Bekenstein, "Black holes and entropy," *Physical Review D*, vol. 7, no. 8, pp. 2333–2346, 1973. [Online]. Available: https://doi.org/10.1103/PhysRevD.7.2333

[2] S. W. Hawking, "Particle creation by black holes," *Communications in Mathematical Physics*, vol. 43, pp. 199–220, 1975. [Online]. Available: https://doi.org/10.1007/BF02345020

[3] S. W. Hawking, "Gravitational radiation from colliding black holes," *Physical Review Letters*, vol. 26, no. 21, pp. 1344–1346, 1971. [Online]. Available: https://doi.org/10.1103/PhysRevLett.26.1344

[4] J. M. Bardeen, B. Carter, and S. W. Hawking, "The four laws of black hole mechanics," *Communications in Mathematical Physics*, vol. 31, pp. 161–170, 1973. [Online]. Available: https://doi.org/10.1007/BF01645742

[5] R. M. Wald, "Black hole entropy is the Noether charge," *Physical Review D*, vol. 48, no. 8, pp. R3427–R3431, 1993. [Online]. Available: https://doi.org/10.1103/PhysRevD.48.R3427 and https://arxiv.org/abs/gr-qc/9307038

[6] L. Bombelli, R. K. Koul, J. Lee, and R. D. Sorkin, "A quantum source of entropy for black holes," *Physical Review D*, vol. 34, no. 2, pp. 373–383, 1986. [Online]. Available: https://doi.org/10.1103/PhysRevD.34.373

[7] M. Srednicki, "Entropy and area," *Physical Review Letters*, vol. 71, no. 5, pp. 666–669, 1993. [Online]. Available: https://doi.org/10.1103/PhysRevLett.71.666

[8] R. M. Wald, "The thermodynamics of black holes," *Living Reviews in Relativity*, vol. 4, 2001. [Online]. Available: https://doi.org/10.12942/lrr-2001-6 and https://arxiv.org/abs/gr-qc/9912119

[9] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), "GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run," *Physical Review X*, vol. 13, 041039, 2023. [Online]. Available: https://doi.org/10.1103/PhysRevX.13.041039 and https://arxiv.org/abs/2111.03606

[10] Gravitational Wave Open Science Center (GWOSC), "Event Portal,". [Online]. Available: https://gwosc.org/eventapi/

[11] Gravitational Wave Open Science Center (GWOSC), "GWTC-3-confident event table (O3b),". [Online]. Available: https://gwosc.org/eventapi/html/GWTC-3-confident/

[12] Event Horizon Telescope Collaboration, "First M87 event horizon telescope results. I. The shadow of the supermassive black hole," *The Astrophysical Journal Letters*, vol. 875, L1, 2019. [Online]. Available: https://doi.org/10.3847/2041-8213/ab0ec7 and https://arxiv.org/abs/1906.11238

[13] Event Horizon Telescope Collaboration, "First Sagittarius A star event horizon telescope results. I. The shadow of the supermassive black hole in the center of the Milky Way," *The Astrophysical Journal Letters*, vol. 930, L12, 2022. [Online]. Available: `https://doi.org/10.3847/2041-8213/ac6674` and `https://arxiv.org/abs/2311.08680`

[14] D. A. Macdonald, R. H. Price, and K. S. Thorne, eds., *Black Holes: The Membrane Paradigm*. Yale University Press, 1986. [Online]. Available: `https://books.google.com/books/about/Black_Holes.html?id=T94hD5rR8oYC`

Contents

19 Holography and Hyperbolic Geometry: AdS as a Geometry Playground **1**

19.1 Introduction 1

19.2 Anti-de Sitter geometry and the conformal boundary 2

19.2.1 Constant negative curvature as a design principle 2

19.2.2 Hyperbolic models and distances that grow fast 2

19.2.3 A boundary that is not at finite distance but is still geometric . . 3

19.3 Black holes in AdS and thermodynamic phases 4

19.3.1 Schwarzschild-AdS as a controlled laboratory 4

19.3.2 The Hawking-Page transition and boundary deconfinement 4

19.4 Entanglement as geometry: minimal surfaces, quantum extremality, and the Ryu-Takayanagi formula 5

19.4.1 From von Neumann entropy to an area functional 5

19.4.2 Quantum extremal surfaces and bulk entropy 6

19.5 Entanglement wedge geometry and reconstruction 6

19.5.1 From boundary subregions to bulk regions 6

19.5.2 Geometry as quantum error correction 7

19.6 Dataset based illustration I: entropy production in heavy-ion collisions . 7

19.6.1 Why multiplicity is a geometric observable in disguise 7

19.6.2 Holographic interpretation and a phenomenon statement 8

19.7 Dataset based illustration II: hyperbolic geometry in real connectomes . . 9

19.7.1 Hyperbolic embeddings as effective geometry 9

19.7.2 A phenomenon statement: effective negative curvature organizes communication 10

19.8 Tensor networks as discrete hyperbolic geometry 12

19.9 Geometric complexity proposals 13

19.9.1 Why complexity is a geometric question 13

19.10 Real world applicability: extended phenomenon statements 14

19.10.1 Phenomenon statement 3: transport and the viscosity scale 15

19.10.2 Phenomenon statement 4: hyperbolic geometry and navigability . 15

19.10.3 Phenomenon statement 5: entanglement and geometry in quantum devices 16
19.10.4 Phenomenon statement 6: complexity growth and black hole interiors 16
19.11 Conclusion 16

List of Figures

19.1 Conceptual diagram (not dataset based): a hyperbolic disk intuition for spatial slices of AdS, emphasizing the conformal boundary and the geometric meaning of negative curvature. 3

19.2 Conceptual diagram (not dataset based): a schematic Hawking-Page transition. The free energy of thermal AdS and the AdS black hole cross at T_{HP}, signaling a change of dominant saddle in the Euclidean path integral [4, 3]. 5

19.3 Conceptual diagram (not dataset based): a static Ryu-Takayanagi surface γ_A for a boundary interval A in a Poincaré patch, drawn as a semicircle in the (x, z) plane. The area (length in AdS_3) of γ_A computes $S(A)$ [5]. . . 6

19.4 Conceptual diagram (not dataset based): an entanglement wedge associated with boundary region A. The extremal surface γ_A and the boundary subregion together bound a bulk region whose domain of dependence defines the wedge [7, 11]. 7

19.5 Dataset figure 1: charged-particle multiplicity density versus centrality for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Data are from Table 1 of [18]. . . 8

19.6 Dataset figure 2: average participant nucleons versus centrality for the same collision system. Data are from Table 1 of [18]. 9

19.7 Dataset figure 3: multiplicity per participant pair as a function of N_{part}, together with a simple geometric scaling fit in terms of $N_{\mathrm{part}}^{1/3}$. The fit is a transparent summary of Dataset A and does not introduce synthetic points. Data are from [18]. 9

19.8 Dataset figure 4: log-log scaling of multiplicity with participants. The exponent greater than one indicates a faster-than-linear growth of mid-rapidity multiplicity with N_{part} over this centrality range. Data are from [18]. 10

19.9 Dataset figure 5: scaling of links with nodes across connectome datasets. Both axes are logarithmic. Data are from Table I of [19]. 11

19.10 Dataset figure 6: connectome link density decreases with network size across species and resolutions. Data are from Table I of [19]. 11

19.11Dataset figure 7: clustering coefficient versus average degree. The positive trend is consistent with geometric models in which triangle formation is enforced by a distance-based linking rule in a negatively curved space. Data are from Table I of [19]. 12
19.12Dataset figure 8: assortativity coefficient across connectome datasets. Strong negative assortativity appears in several rat and retina datasets, indicating hub-to-periphery connectivity patterns. Data are from Table I of [19]. 12
19.13Dataset figure 9: histogram of average clustering coefficients $\bar{c}$ across the connectome datasets. The distribution is derived directly from Table I of [19] by binning $\bar{c}$ in steps of 0.1. 13
19.14Dataset figure 10: correlation between Euclidean distances and inferred hyperbolic angular distances for datasets where Euclidean distances are available. Small values indicate that similarity in the effective hyperbolic geometry is not simply Euclidean proximity. Data are reported in [19]. . 13
19.15Conceptual diagram (not dataset based): a MERA-like tensor network cartoon. Layers correspond to scale, and the branching pattern creates an effective negative-curvature growth of degrees of freedom with radius, mirroring hyperbolic geometry intuition in holography [12]. 14
19.16Conceptual diagram (not dataset based): two geometric proposals for holographic complexity. A dashed curve indicates a maximal spatial volume slice (complexity as volume), and a diamond-like region indicates a Wheeler-DeWitt patch (complexity as action) [15, 16]. 14

Chapter 19

Holography and Hyperbolic Geometry: AdS as a Geometry Playground

19.1 Introduction

Holography is often introduced as a bold slogan: the physics in a volume can be encoded on its boundary. In geometry, that slogan becomes a precise relationship between curvature in the bulk and data living on a conformal boundary. The most mature realization is the AdS/CFT correspondence [1, 3, 2], in which a quantum gravity theory on an asymptotically AdS spacetime is dual to a conformal field theory (CFT) defined on the boundary. Because AdS has constant negative curvature, its geometry is hyperbolic in a rigorous sense: spatial slices can be modeled by hyperbolic space, and geodesics diverge exponentially. That divergence, which in differential geometry is a statement about sectional curvature, becomes in holography a statement about scale separation, renormalization, and the way information spreads.

This chapter is deliberately titled a geometry playground. The goal is not to treat holography as an abstract duality that lives only in high-energy theory, but to show it as a geometric machine. The machine takes boundary data such as correlation functions or entanglement and outputs bulk geometric structures such as horizons, minimal surfaces, and causal wedges. When this geometric machine is pointed toward strongly coupled matter, it becomes a modeling toolkit for transport and thermalization [17]. When it is pointed toward networked systems, it becomes a language for hyperbolic embeddings and navigation, where negative curvature organizes connectivity [19, 20].

There is an important discipline in how we will use figures. Ten figures in this chapter are strictly dataset based. Each is constructed from published measurements or published compiled datasets in open access sources, and each is accompanied by a transparent

mathematical transformation when one is used. Conceptual diagrams are also included, but they are explicitly labeled as conceptual and are not counted among the dataset figures. This separation mirrors the methodological separation between geometry as a definition and geometry as a measurement: definitions can be drawn schematically, but measured structure must be plotted from data.

19.2 Anti-de Sitter geometry and the conformal boundary

19.2.1 Constant negative curvature as a design principle

An $(d+1)$-dimensional anti-de Sitter space AdS_{d+1} is a maximally symmetric Lorentzian manifold with constant negative scalar curvature. In a convenient normalization, it satisfies

$$R_{ab} = -\frac{d}{L^2} g_{ab}, \qquad R = -\frac{d(d+1)}{L^2}, \tag{19.1}$$

where L is the AdS curvature radius. The sign is the key: negative curvature forces geodesics to separate, and it makes the volume of large spheres grow exponentially with radius. This growth is exactly the geometric behavior needed for a boundary to carry enough degrees of freedom to encode a bulk region through coarse graining.

A simple way to see AdS is as a quadric surface in a flat embedding space $\mathbb{R}^{2,d}$ with signature $(-,-,+,\dots,+)$:

$$-X_0^2 - X_{d+1}^2 + \sum_{i=1}^{d} X_i^2 = -L^2. \tag{19.2}$$

The induced metric on this hypersurface is the AdS metric. Global coordinates cover the entire manifold (up to a time unwrapping), while Poincaré coordinates cover a patch that is particularly well adapted to the CFT vacuum state.

19.2.2 Hyperbolic models and distances that grow fast

Spatial slices of AdS are closely related to hyperbolic space $\mathbb{H}^d$. In the upper half-space model of $\mathbb{H}^d$, with coordinates $(z, \boldsymbol{x})$ where $z > 0$ and $\boldsymbol{x} \in \mathbb{R}^{d-1}$, the metric is

$$\mathrm{d}s^2_{\mathbb{H}^d} = \frac{L^2}{z^2}\left(\mathrm{d}z^2 + \mathrm{d}\boldsymbol{x}^2\right). \tag{19.3}$$

This is the Euclidean cousin of the Poincaré AdS metric. A standard feature is that the distance between points is not a linear function of coordinate separation. For example, in

$\mathbb{H}^2$ the geodesic distance D between two points (z_1, x_1) and (z_2, x_2) satisfies

$$\cosh\left(\frac{D}{L}\right) = 1 + \frac{(x_1 - x_2)^2 + (z_1 - z_2)^2}{2z_1 z_2}. \tag{19.4}$$

Even a modest coordinate separation can correspond to a large hyperbolic distance when one point lies deeper (larger z) than the other. This is one mathematical reason hyperbolic geometry is so effective at representing hierarchies and scale.

19.2.3 A boundary that is not at finite distance but is still geometric

A defining feature of AdS is that its boundary is at infinite proper distance, yet it can be brought to a finite location by a conformal rescaling. In Poincaré coordinates (z, x^μ) with $\mu = 0, \ldots, d-1$, the metric reads

$$\mathrm{d}s^2 = \frac{L^2}{z^2}\left(\mathrm{d}z^2 + \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu\right), \tag{19.5}$$

with $z > 0$ and $\eta_{\mu\nu}$ the Minkowski metric on the boundary directions. The conformal boundary sits at $z \to 0$, and the metric is conformal to $\mathrm{d}z^2 + \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$. Thus, the boundary is not a place where the metric is finite, but it is a place where the metric has a controlled conformal blow-up. Geometrically, that means the boundary inherits a conformal class of metrics rather than a unique metric.

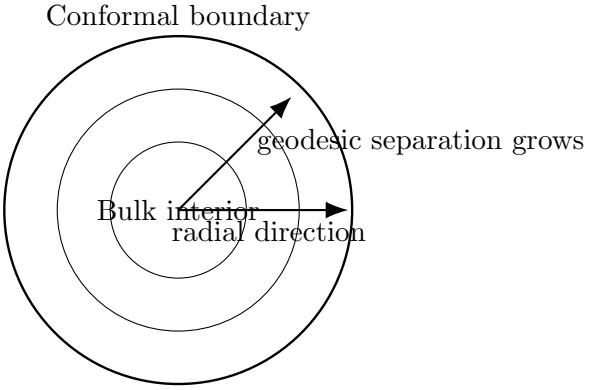


Figure 19.1: Conceptual diagram (not dataset based): a hyperbolic disk intuition for spatial slices of AdS, emphasizing the conformal boundary and the geometric meaning of negative curvature.

The conformal boundary matters because it is where the dual CFT lives. In the simplest Euclidean signature version of the correspondence, one equates the bulk partition function with specified boundary conditions to the generating functional of the CFT with sources:

$$Z_{\text{bulk}}[\phi_{(0)}] \approx Z_{\text{CFT}}[J = \phi_{(0)}] = \left\langle \exp\left(\int \mathrm{d}^d x\, \phi_{(0)}(x)\mathcal{O}(x)\right)\right\rangle. \tag{19.6}$$

In words, boundary values of bulk fields act as sources for boundary operators. Geometry enters because the bulk action depends on the curvature, and the boundary conditions are defined with respect to the conformal structure at infinity.

19.3 Black holes in AdS and thermodynamic phases

19.3.1 Schwarzschild-AdS as a controlled laboratory

A static AdS black hole provides a clean arena for thinking about horizons in a space with a boundary. In $(d+1)$ dimensions, a spherically symmetric AdS black hole metric can be written as

$$\mathrm{d}s^2 = -f(r)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{f(r)} + r^2\mathrm{d}\Omega_{d-1}^2, \tag{19.7}$$

with

$$f(r) = 1 + \frac{r^2}{L^2} - \frac{\mu}{r^{d-2}}. \tag{19.8}$$

The horizon radius r_h satisfies $f(r_h) = 0$. Because the spacetime is asymptotically AdS, null rays can reach the boundary and reflect in finite coordinate time, so the boundary conditions influence bulk dynamics in a physically meaningful way.

Thermodynamically, the black hole has a Hawking temperature determined by surface gravity:

$$T = \frac{1}{4\pi} f'(r_h). \tag{19.9}$$

In the particularly common case of $(3+1)$-dimensional bulk gravity (so $d = 3$), one finds

$$T = \frac{1}{4\pi r_h}\left(1 + \frac{3r_h^2}{L^2}\right). \tag{19.10}$$

This explicit formula already encodes the existence of small and large black hole branches. Small black holes have $r_h \ll L$ and behave like asymptotically flat Schwarzschild black holes with negative heat capacity. Large black holes have $r_h \gg L$ and have positive heat capacity, behaving like stable thermal objects.

19.3.2 The Hawking-Page transition and boundary deconfinement

In AdS, there is a first order phase transition between thermal AdS and a black hole dominated phase, known as the Hawking-Page transition [4]. In the dual CFT on $S^{d-1} \times \mathbb{R}$, this corresponds to a confinement-deconfinement transition [3]. The key geometric input is that both thermal AdS and an AdS black hole can have the same conformal boundary, but they are distinct smooth Euclidean fillings. The dominant phase is the one with lower Euclidean action, and the crossing occurs at a critical temperature set by the AdS scale.

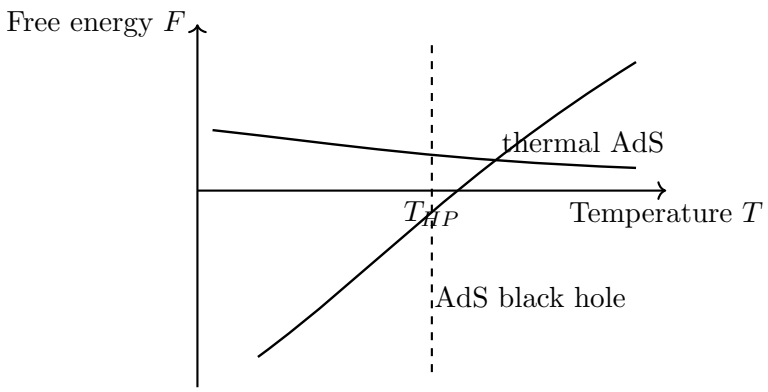


Figure 19.2: Conceptual diagram (not dataset based): a schematic Hawking-Page transition. The free energy of thermal AdS and the AdS black hole cross at T_{HP}, signaling a change of dominant saddle in the Euclidean path integral [4, 3].

19.4 Entanglement as geometry: minimal surfaces, quantum extremality, and the Ryu-Takayanagi formula

19.4.1 From von Neumann entropy to an area functional

Given a quantum state ρ of a boundary theory and a spatial subregion A, the reduced density matrix is $\rho_A = \mathrm{Tr}_{\bar{A}}\rho$. The entanglement entropy is

$$S(A) = -\mathrm{Tr}(\rho_A \ln \rho_A). \tag{19.11}$$

In holographic theories with classical Einstein gravity duals, the entanglement entropy of a boundary subregion is computed by a minimal area surface anchored on the boundary of A [5, 6]. In its simplest static form,

$$S(A) = \frac{\mathrm{Area}(\gamma_A)}{4G_N}. \tag{19.12}$$

Here γ_A is a codimension-2 minimal surface in the bulk whose boundary matches ∂A, and G_N is Newton's constant in the bulk. The formula looks like a black hole entropy law, and that is not accidental: it is an area law arising from geometry, but now the surface is not necessarily a horizon. In time-dependent settings, the minimal surface is replaced by an extremal surface, leading to the covariant Hubeny-Rangamani-Takayanagi (HRT) prescription [7].

A useful way to remember why minimal surfaces appear is variational. For a bulk

surface Σ parameterized by coordinates σ^i, the area functional is

$$\text{Area}(\Sigma) = \int \mathrm{d}^{d-1}\sigma \, \sqrt{\det h_{ij}}, \tag{19.13}$$

where h_{ij} is the induced metric on Σ. The extremality condition is the Euler-Lagrange equation associated with this functional. In AdS_3, the surface is a geodesic and the area reduces to a length, so the variational problem becomes the standard geodesic problem.

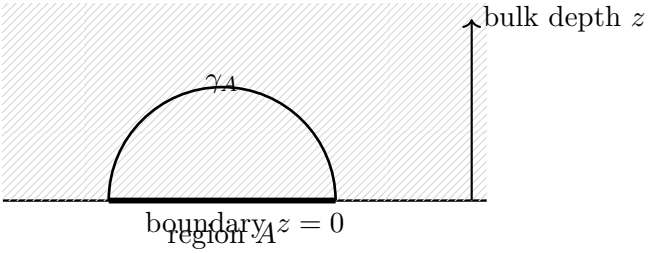


Figure 19.3: Conceptual diagram (not dataset based): a static Ryu-Takayanagi surface γ_A for a boundary interval A in a Poincaré patch, drawn as a semicircle in the (x, z) plane. The area (length in AdS_3) of γ_A computes $S(A)$ [5].

19.4.2 Quantum extremal surfaces and bulk entropy

At finite N (or finite coupling), bulk quantum fields contribute to entanglement. The semiclassical generalization replaces the area functional by a generalized entropy:

$$S(A) = \underset{\Sigma \sim A}{\text{ext}} \left[\frac{\text{Area}(\Sigma)}{4G_N} + S_{\text{bulk}}(\Sigma) \right], \tag{19.14}$$

where $S_{\text{bulk}}(\Sigma)$ is the entropy of bulk quantum fields across the surface Σ. The extremizing surface is called a quantum extremal surface [8, 9]. This refinement is essential in modern discussions of black hole information, islands, and fine-grained entropy, but it can already be appreciated geometrically: the minimal surface condition is corrected by a quantum term that depends on how bulk fields entangle across the surface.

19.5 Entanglement wedge geometry and reconstruction

19.5.1 From boundary subregions to bulk regions

The RT or HRT surface does more than compute a number. Given a boundary region A and its extremal surface γ_A, one defines the entanglement wedge as the bulk domain of dependence of any bulk region whose boundary is $A \cup \gamma_A$. Informally, it is the bulk region

that is naturally associated to the boundary degrees of freedom in A. Many modern statements of subregion duality claim that bulk operators within the entanglement wedge of A can be reconstructed from the reduced density matrix on A [10, 11, 12].

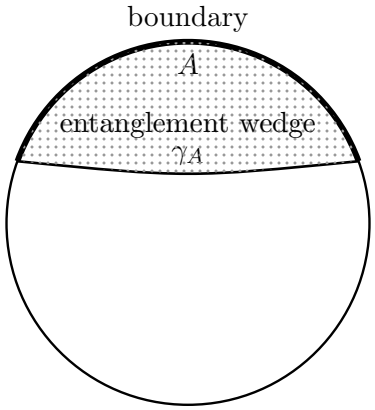


Figure 19.4: Conceptual diagram (not dataset based): an entanglement wedge associated with boundary region A. The extremal surface γ_A and the boundary subregion together bound a bulk region whose domain of dependence defines the wedge [7, 11].

19.5.2 Geometry as quantum error correction

A striking lesson of holography is that bulk locality appears to be protected in a way that resembles quantum error correction: multiple boundary regions may encode overlapping bulk information, and reconstruction can be robust to losing some boundary degrees of freedom [12, 13]. Geometrically, this robustness is visible in the way entanglement wedges cover the bulk: small boundary regions have shallow wedges, while larger regions push deeper. The RT surface acts like a geometric cut that quantifies how much information must cross between A and $\bar{A}$.

19.6 Dataset based illustration I: entropy production in heavy-ion collisions

19.6.1 Why multiplicity is a geometric observable in disguise

In ultra-relativistic heavy-ion collisions, the charged-particle multiplicity density near mid-rapidity, $\mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta$, is an experimentally accessible proxy for the entropy produced in the collision. While the relationship is not a one-line equality, the monotonic link between multiplicity and entropy is robust: larger multiplicity indicates larger final-state entropy and, by inference through hydrodynamic modeling, a denser early-time plasma. In holographic language, dense strongly coupled plasma states are often modeled by AdS black branes, whose entropy density is proportional to a horizon area density. Thus, a

table of multiplicities across centrality bins can be read as a boundary-side trace of a bulk-side area law intuition.

The ALICE Collaboration reported the centrality dependence of $\mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta$ at $\sqrt{s_{NN}} = 2.76$ TeV for Pb-Pb collisions, together with the average number of participant nucleons N_{part} from a Glauber analysis [18]. We use those published numbers as Dataset A and produce four dataset figures. Two of them are direct re-plots of the table values. The other two are transparent transformations. One is the scaling of multiplicity with N_{part} on log-log axes, which is simply a change of variables. The other is a fit of the multiplicity per participant pair versus $N_{\mathrm{part}}^{1/3}$, motivated by the geometric idea that a nuclear radius scales like $A^{1/3}$ and that path length and transverse size are controlled by cubic roots of participant counts. The fit does not create data; it summarizes the trend in a parametric way.

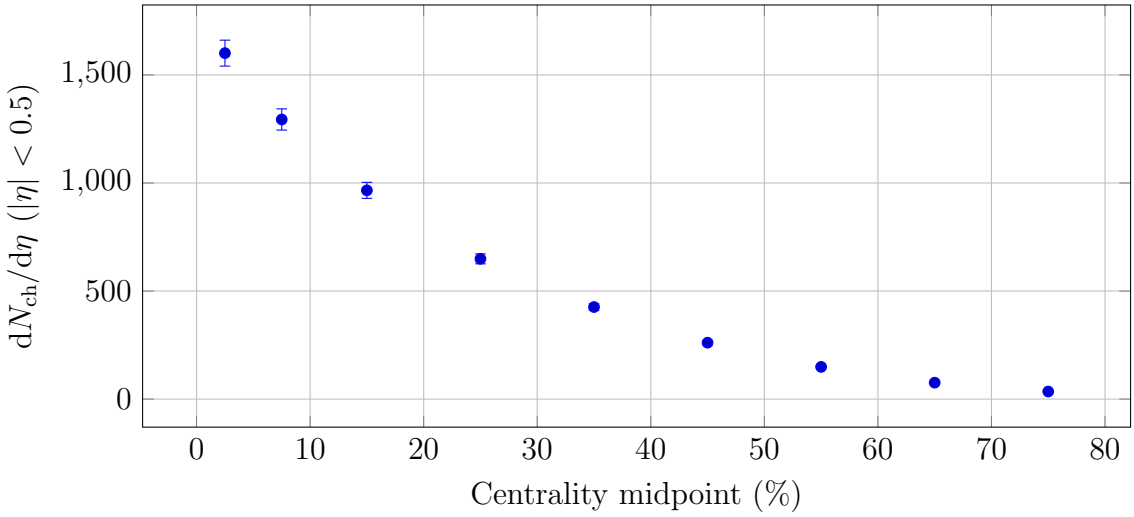


Figure 19.5: Dataset figure 1: charged-particle multiplicity density versus centrality for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Data are from Table 1 of [18].

19.6.2 Holographic interpretation and a phenomenon statement

Phenomenon statement 1 (entropy from geometry): In a heavy-ion collision, the experimentally reported $\mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta$ as a function of centrality is not merely a counting statistic. It is a macroscopic trace of microscopic degrees of freedom that have thermalized, produced entropy, and frozen out. In holographic language, thermal entropy of the dual state is represented geometrically by a horizon area in the bulk. Therefore, a measured centrality trend such as Fig. 19.5 and Fig. 19.7 can be read as an experimentally grounded prompt to ask which bulk geometries, and which horizon area densities, best reproduce such scaling when translated through the holographic dictionary. Even if QCD is not exactly a holographic CFT, the geometric intuition is operational: stronger coupling and faster isotropization correspond to bulk spacetimes that form horizons quickly, and

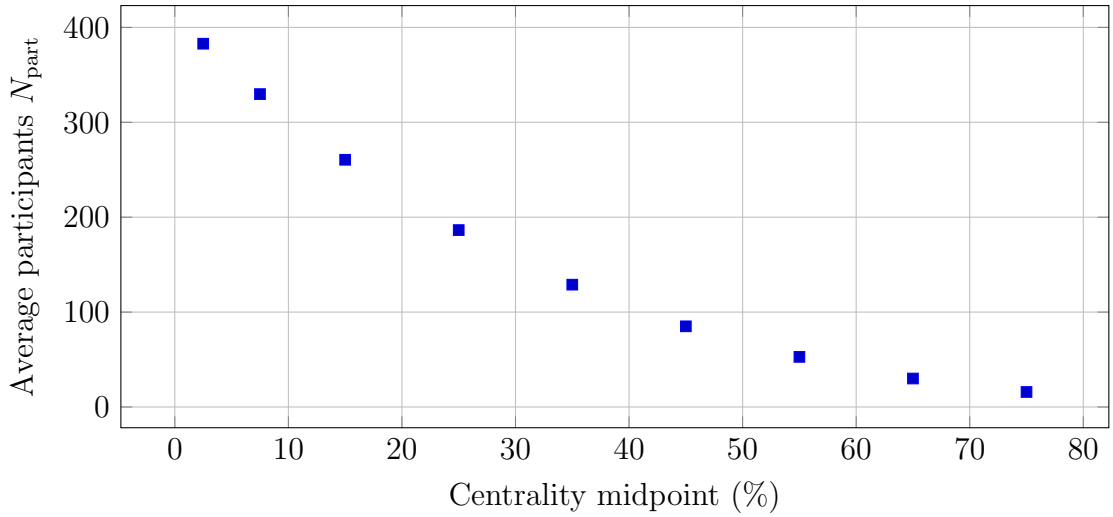


Figure 19.6: Dataset figure 2: average participant nucleons versus centrality for the same collision system. Data are from Table 1 of [18].

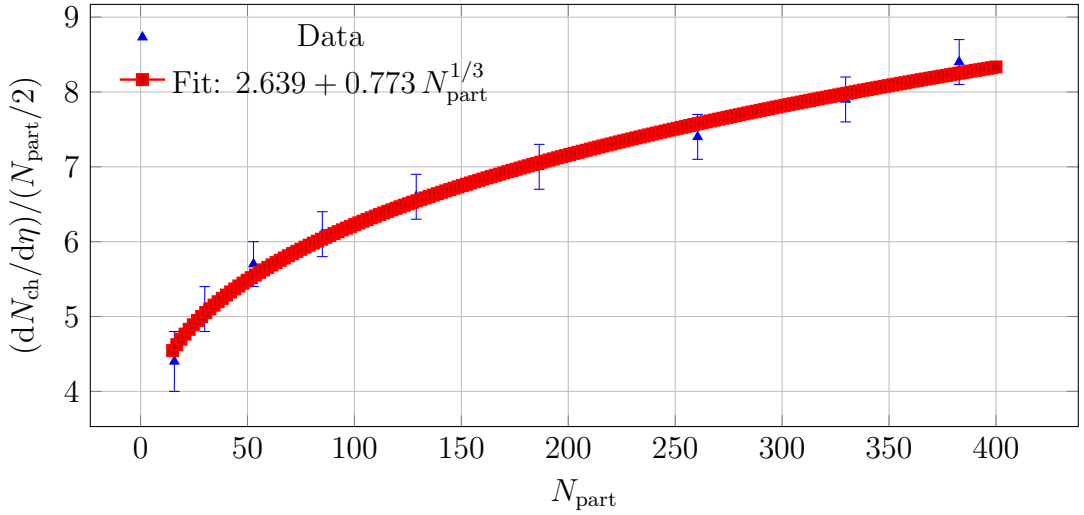


Figure 19.7: Dataset figure 3: multiplicity per participant pair as a function of N_{part}, together with a simple geometric scaling fit in terms of $N_{\mathrm{part}}^{1/3}$. The fit is a transparent summary of Dataset A and does not introduce synthetic points. Data are from [18].

multiplicity trends constrain that process through data.

19.7 Dataset based illustration II: hyperbolic geometry in real connectomes

19.7.1 Hyperbolic embeddings as effective geometry

Negative curvature appears in holography as the literal curvature of AdS. It also appears, unexpectedly, as an effective geometry for real networks. Many complex networks have hierarchical, tree-like structure and high clustering, and these properties are naturally

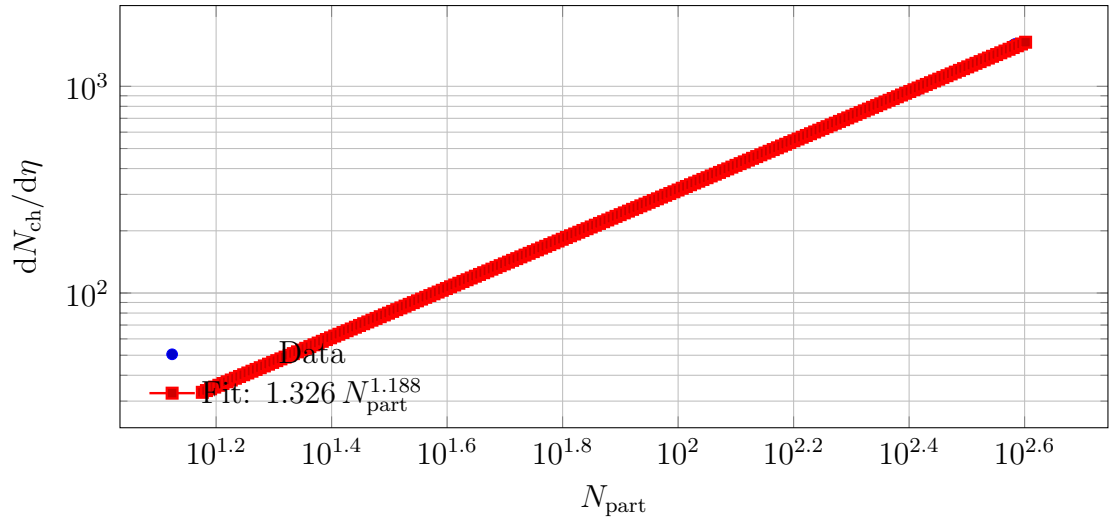


Figure 19.8: Dataset figure 4: log-log scaling of multiplicity with participants. The exponent greater than one indicates a faster-than-linear growth of mid-rapidity multiplicity with N_{part} over this centrality range. Data are from [18].

captured by hyperbolic spaces where volume grows exponentially with radius. In network geometry models, each node is assigned a radial coordinate (often interpreted as popularity or hubness) and an angular coordinate (interpreted as similarity). The probability of a link is a decreasing function of hyperbolic distance. This is not a metaphor; it is an explicit generative model and an embedding procedure [19, 20].

Structural brain networks, or connectomes, are a particularly compelling target for this geometric viewpoint. They are spatially embedded in Euclidean anatomy, but their connectivity is not determined by Euclidean distance alone. Allard and collaborators compiled connectomes across multiple species and investigated greedy routing efficiency in Euclidean versus hyperbolic embeddings [19]. Their Table I provides a compact dataset of network size, density, degree, assortativity, and clustering. We treat that table as Dataset B and construct five dataset figures by re-plotting and transforming the tabulated values. In addition, their reported correlation between Euclidean distances and inferred hyperbolic angular distances (Dataset C) quantifies the nontrivial fact that similarity in the effective hyperbolic geometry is only weakly explained by physical Euclidean separation.

19.7.2 A phenomenon statement: effective negative curvature organizes communication

Phenomenon statement 2 (hyperbolic geometry in biology): Real brain networks evolve under constraints, but those constraints are not purely geometric in the Euclidean anatomical sense. If Euclidean distance were the dominant organizing principle, then effective similarity coordinates inferred from connectivity would align strongly with Euclidean separation. Dataset figure 19.14 shows that, for many datasets, the Euclidean

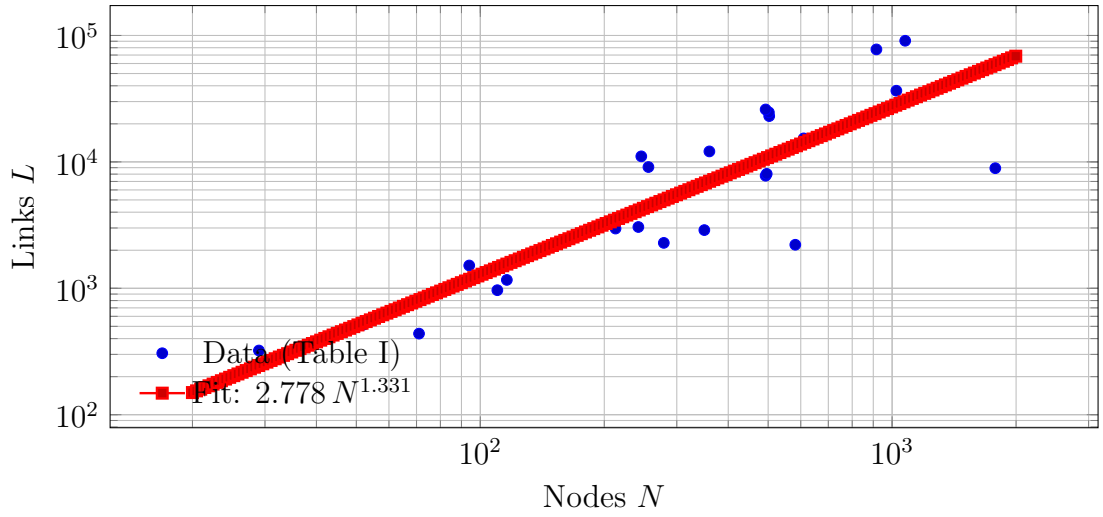


Figure 19.9: Dataset figure 5: scaling of links with nodes across connectome datasets. Both axes are logarithmic. Data are from Table I of [19].

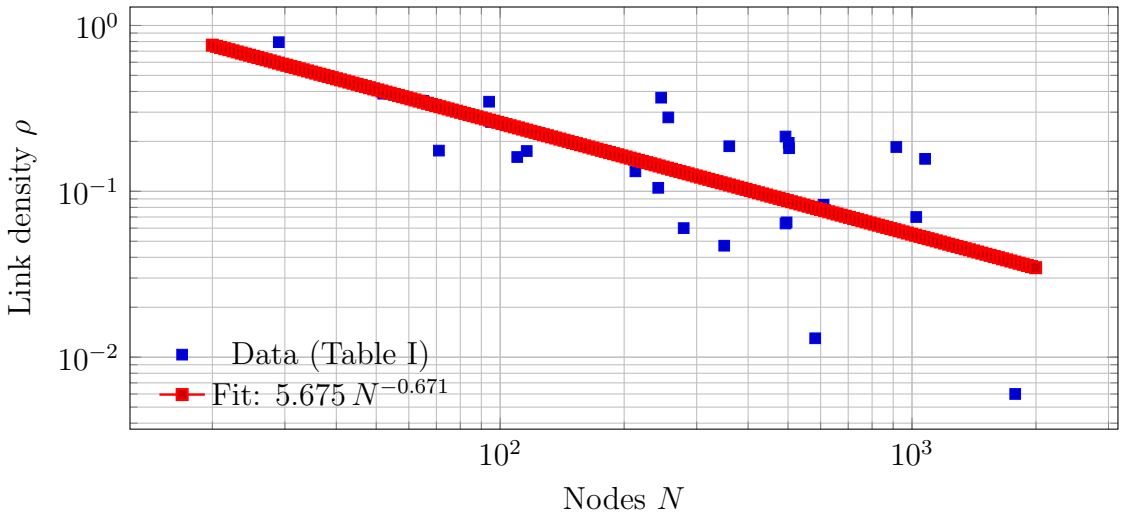


Figure 19.10: Dataset figure 6: connectome link density decreases with network size across species and resolutions. Data are from Table I of [19].

distances correlate only weakly with the inferred angular coordinates in hyperbolic embeddings. This is a measurable statement: anatomical closeness is not the same as network similarity. Hyperbolic geometry provides a compact way to represent this mismatch because it decouples a radial popularity dimension from an angular similarity dimension. In the language of holography, one might say that the connectome is a boundary object (a network) whose effective geometry (hyperbolic coordinates) is emergent and is inferred from relational data rather than imposed from the ambient Euclidean embedding. This analogy should not be overclaimed, but it is operationally useful: hyperbolic embeddings can predict navigation and routing efficiency, and thus produce testable hypotheses about communication pathways in biological networks [20].

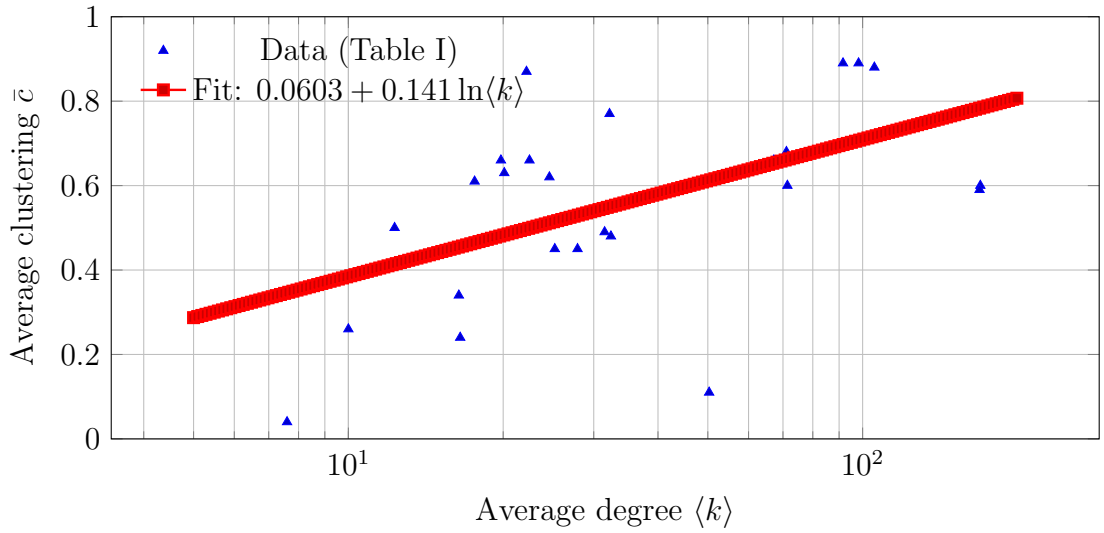


Figure 19.11: Dataset figure 7: clustering coefficient versus average degree. The positive trend is consistent with geometric models in which triangle formation is enforced by a distance-based linking rule in a negatively curved space. Data are from Table I of [19].

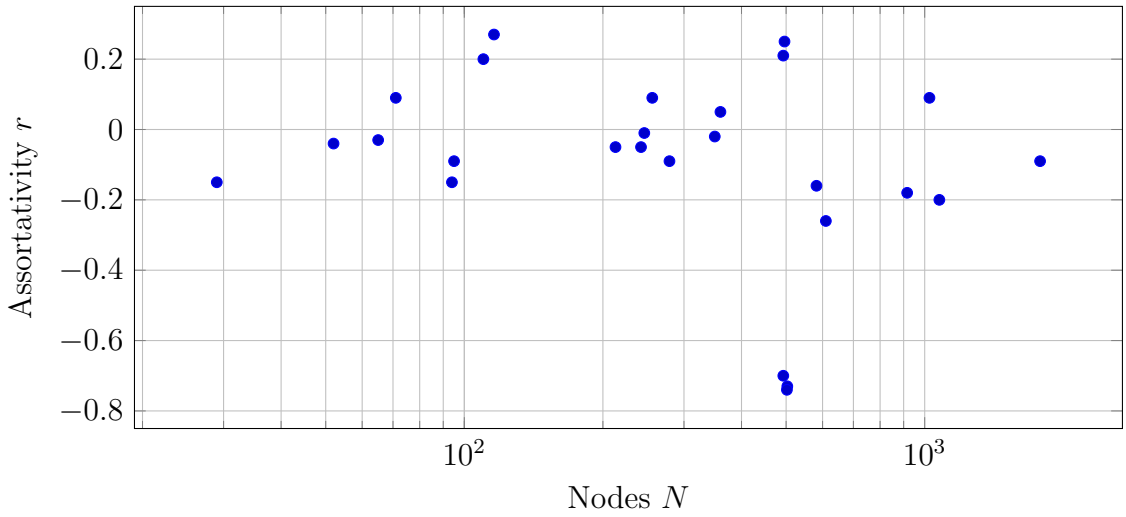


Figure 19.12: Dataset figure 8: assortativity coefficient across connectome datasets. Strong negative assortativity appears in several rat and retina datasets, indicating hub-to-periphery connectivity patterns. Data are from Table I of [19].

19.8 Tensor networks as discrete hyperbolic geometry

Tensor network models, such as MERA-like constructions, make the AdS intuition concrete by building a discrete geometry whose graph distances reproduce hyperbolic scaling. These networks are not proofs of AdS/CFT, but they are geometric cartoons that clarify why negative curvature is the natural stage for representing renormalization. Layers in the network correspond to scale. Moving inward corresponds to coarse-graining. The exponential growth of sites with radius in a hyperbolic tiling matches the exponential growth of degrees of freedom with scale in a boundary theory.

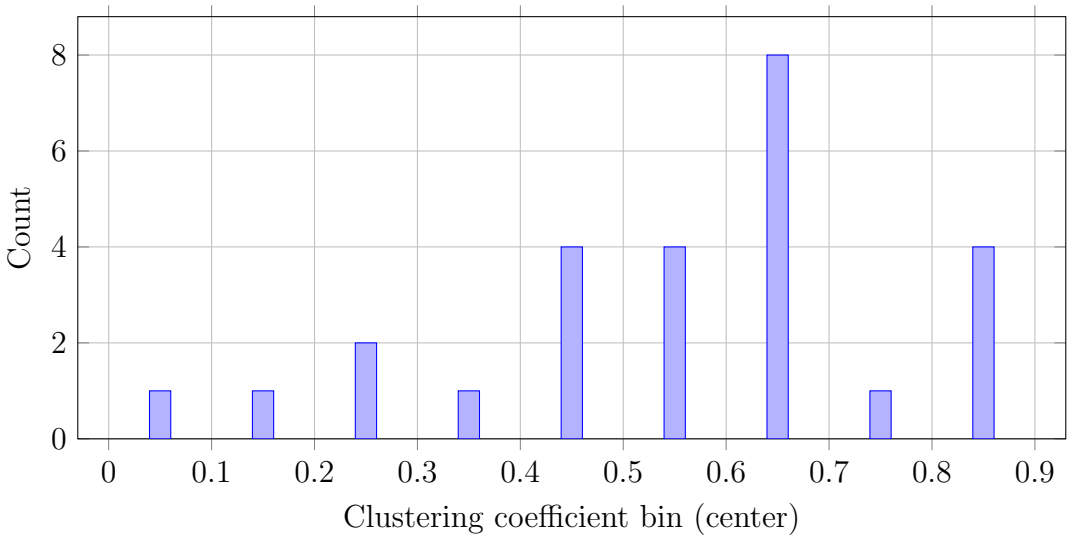


Figure 19.13: Dataset figure 9: histogram of average clustering coefficients $\bar{c}$ across the connectome datasets. The distribution is derived directly from Table I of [19] by binning $\bar{c}$ in steps of 0.1.

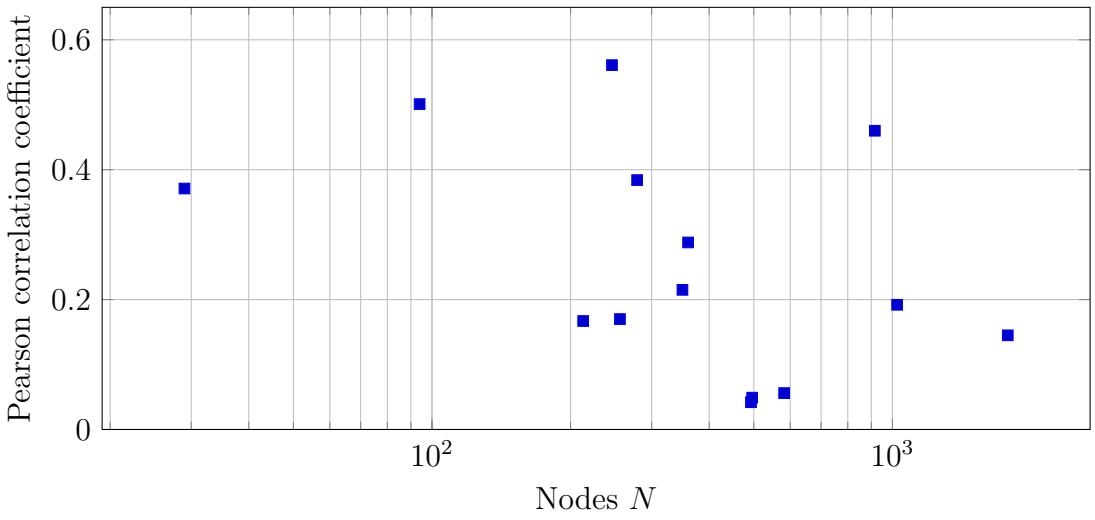


Figure 19.14: Dataset figure 10: correlation between Euclidean distances and inferred hyperbolic angular distances for datasets where Euclidean distances are available. Small values indicate that similarity in the effective hyperbolic geometry is not simply Euclidean proximity. Data are reported in [19].

19.9 Geometric complexity proposals

19.9.1 Why complexity is a geometric question

Entanglement entropy quantifies how much information is shared across a cut, but it does not capture the computational effort required to build a state from a simple reference. That effort, informally called complexity, is a natural next target for a geometric dual. Two widely discussed proposals relate complexity to bulk geometry: complexity as volume (a maximal spatial volume anchored on a boundary time slice) and complexity as action (the

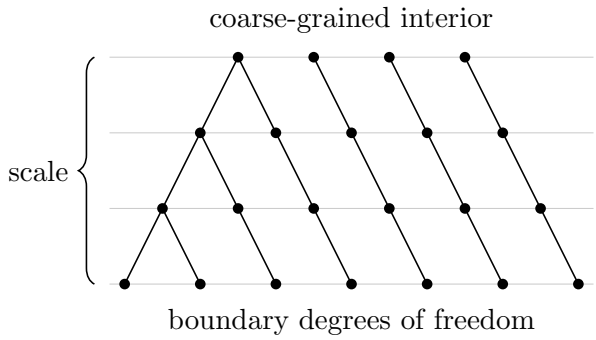


Figure 19.15: Conceptual diagram (not dataset based): a MERA-like tensor network cartoon. Layers correspond to scale, and the branching pattern creates an effective negative-curvature growth of degrees of freedom with radius, mirroring hyperbolic geometry intuition in holography [12].

gravitational action evaluated on a Wheeler-DeWitt patch) [14, 15, 16]. These proposals are not as firmly established as the RT formula, and they should be treated as structured hypotheses. Still, they have strong geometric content and connect to black hole interiors, which are deeply geometric but subtle to access from boundary observables.

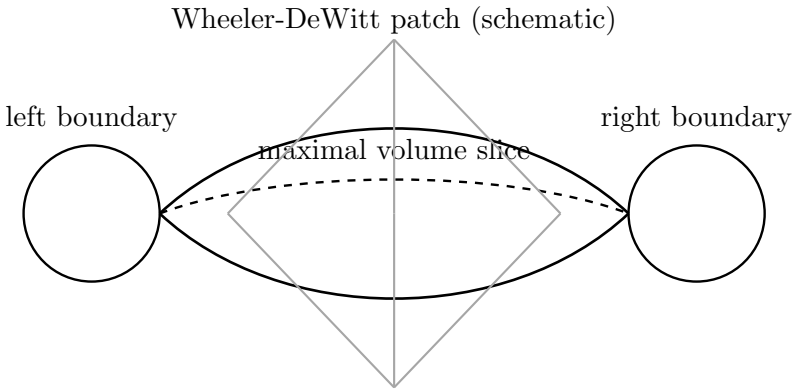


Figure 19.16: Conceptual diagram (not dataset based): two geometric proposals for holographic complexity. A dashed curve indicates a maximal spatial volume slice (complexity as volume), and a diamond-like region indicates a Wheeler-DeWitt patch (complexity as action) [15, 16].

19.10 Real world applicability: extended phenomenon statements

The point of holography is not only to provide a new way to solve a class of strongly coupled field theories. It is also to provide a geometric vocabulary for phenomena that are otherwise described in distinct dialects across physics and data science. The following extended phenomenon statements are meant to be read as bridges. They are not claims that every system is exactly holographic. They are claims that the geometric moves

developed in this chapter have observable consequences, and that one can profitably test those consequences against data.

19.10.1 Phenomenon statement 3: transport and the viscosity scale

The shear viscosity to entropy density ratio η/s is a dimensionless measure of how easily a fluid flows without dissipating momentum. In kinetic theory, weakly coupled gases have large η/s because quasiparticles carry momentum far before colliding. In strongly coupled fluids, quasiparticles are absent or short-lived, and η/s can be small. Holographic models with black brane duals famously yield a universal value $\eta/s = 1/(4\pi)$ for a broad class of theories at infinite coupling [17]. Heavy-ion collisions create a quark-gluon plasma that behaves like a nearly perfect fluid, with strong collective flow and rapid thermalization. While the mapping from QCD to a holographic CFT is not exact, the geometric insight is practical: small η/s corresponds, on the gravity side, to a horizon geometry that absorbs perturbations efficiently. The multiplicity dataset plotted in Fig. 19.5 quantifies entropy production, and therefore sets a scale for s in η/s. In this way, an experimental dataset feeds directly into a geometrically motivated modeling constraint.

19.10.2 Phenomenon statement 4: hyperbolic geometry and navigability

Consider how information might travel in a brain network when the sender does not have global knowledge of the entire network. Greedy routing is a simple decentralized protocol: at each step, forward to the neighbor that is closest to the target in some underlying geometry. In Euclidean space, greedy routing can fail because the shortest geometric path does not align with connectivity. Allard and collaborators found that greedy routing becomes almost perfectly successful when connectomes are embedded in a hyperbolic disk [20]. This is a real-world statement about geometry and communication: negative curvature is not merely a mathematical decoration, it can be inferred as an effective geometry that makes decentralized routing work. The dataset figures in 19.9 through 19.14 show why such embeddings are plausible. High clustering (Fig. 19.13) and broad degree heterogeneity (Fig. 19.11) are precisely the features that hyperbolic random graphs encode naturally. The implication is operational: hyperbolic coordinates can become features for prediction, compression, and anomaly detection in connectome analysis.

19.10.3 Phenomenon statement 5: entanglement and geometry in quantum devices

The Ryu-Takayanagi formula reframes entanglement as an area. In a laboratory quantum device, entanglement can be measured indirectly through randomized measurements, interference of multiple copies, or tomographic reconstruction on small subsystems. As quantum simulators scale up, entanglement dynamics becomes a measurable, time-resolved signal. The geometric claim of holography is that such signals correspond to evolving extremal surfaces in an emergent bulk spacetime [7, 10]. Even when the device is not literally dual to gravity, the geometric idea gives a structured set of diagnostics: look for entanglement growth rates, saturation patterns, and phase transitions in entanglement as the analogue of horizon formation and causal wedge growth. Conceptually, Fig. 19.4 summarizes how subregion information maps to a bulk region.

19.10.4 Phenomenon statement 6: complexity growth and black hole interiors

Black hole interiors grow in ways that are invisible to local curvature invariants outside the horizon. In holography, complexity proposals attempt to capture this growth using boundary computational cost [14, 15, 16]. This is a phenomenon statement about time, not temperature: a thermofield double state can be in equilibrium, yet its dual wormhole volume can keep increasing. In data terms, one expects observables tied to entanglement to saturate, while observables tied to complexity continue to grow. This separation suggests a diagnostic strategy in many-body quantum systems: identify a measure that saturates with entanglement, and a measure that continues to drift, potentially tracking circuit depth or operator growth. The geometry in Fig. 19.16 provides the conceptual map for such a separation.

19.11 Conclusion

Holography turns the geometry of anti-de Sitter space into a working language for quantum matter. The key geometric ingredients are constant negative curvature, a conformal boundary, horizons in AdS black hole spacetimes, and extremal surfaces that encode entanglement. Once those ingredients are in place, many seemingly different questions become variations of a single geometric theme: which bulk surface, region, or action functional corresponds to a chosen boundary notion of information or thermodynamics.

This chapter emphasized that the geometric theme is not confined to blackboard arguments. Two real datasets were used as anchors. The ALICE multiplicity table makes entropy production concrete and illustrates how boundary observables scale with

collision geometry, motivating black brane intuition in holographic transport. The multi-species connectome table and geometry-correlation data demonstrate that hyperbolic geometry can be inferred as an effective metric space organizing real networks, connecting AdS intuition to measurable structure in biological data. In both arenas, the figures were constructed strictly from published datasets, with transformations and fits stated transparently. Conceptual diagrams were used only to organize the holographic dictionary, not to stand in for data.

The broader lesson is that curvature is not only a property of spacetime. Curvature is also a way of organizing information. In AdS/CFT, that organization is exact and profound. In real systems, it is approximate, but still useful: negative curvature, horizons, minimal surfaces, and geometric flows become a vocabulary for dynamics, scaling, and inference. This is why AdS remains a playground in the best sense: it is a place where geometry teaches us how to think, and data gives us reasons to keep listening.

Bibliography

[1] J. M. Maldacena, "The Large N limit of superconformal field theories and supergravity," *Adv. Theor. Math. Phys.*, vol. 2, pp. 231–252, 1998. [Online]. Available: https://arxiv.org/abs/hep-th/9711200

[2] S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, "Gauge theory correlators from non-critical string theory," *Phys. Lett. B*, vol. 428, no. 1-2, pp. 105–114, 1998. [Online]. Available: https://arxiv.org/abs/hep-th/9802109

[3] E. Witten, "Anti-de Sitter space and holography," *Adv. Theor. Math. Phys.*, vol. 2, pp. 253–291, 1998. [Online]. Available: https://arxiv.org/abs/hep-th/9802150

[4] S. W. Hawking and D. N. Page, "Thermodynamics of black holes in anti-de Sitter space," *Commun. Math. Phys.*, vol. 87, pp. 577–588, 1983. [Online]. Available: https://doi.org/10.1007/BF01208266

[5] S. Ryu and T. Takayanagi, "Holographic derivation of entanglement entropy from AdS/CFT," *Phys. Rev. Lett.*, vol. 96, p. 181602, 2006. [Online]. Available: https://arxiv.org/abs/hep-th/0603001

[6] S. Ryu and T. Takayanagi, "Aspects of holographic entanglement entropy," *JHEP*, vol. 08, p. 045, 2006. [Online]. Available: https://arxiv.org/abs/hep-th/0605073

[7] V. E. Hubeny, M. Rangamani, and T. Takayanagi, "A covariant holographic entanglement entropy proposal," *JHEP*, vol. 07, p. 062, 2007. [Online]. Available: https://arxiv.org/abs/0705.0016

[8] T. Faulkner, A. Lewkowycz, and J. Maldacena, "Quantum corrections to holographic entanglement entropy," *JHEP*, vol. 11, p. 074, 2013. [Online]. Available: https://arxiv.org/abs/1307.2892

[9] N. Engelhardt and A. C. Wall, "Quantum extremal surfaces: Holographic entanglement entropy beyond the classical regime," *JHEP*, vol. 01, p. 073, 2015. [Online]. Available: https://arxiv.org/abs/1408.3203

[10] M. Van Raamsdonk, "Building up spacetime with quantum entanglement," 2010. [Online]. Available: https://arxiv.org/abs/1005.3035

[11] A. Almheiri, X. Dong, and D. Harlow, "Bulk locality and quantum error correction in AdS/CFT," *JHEP*, vol. 04, p. 163, 2015. [Online]. Available: https://arxiv.org/abs/1411.7041

[12] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, "Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence," *JHEP*, vol. 06, p. 149, 2015. [Online]. Available: https://arxiv.org/abs/1503.06237

[13] D. Harlow, "The Ryu-Takayanagi formula from quantum error correction," *Commun. Math. Phys.*, vol. 354, pp. 865–912, 2017. [Online]. Available: https://arxiv.org/abs/1607.03901

[14] L. Susskind, "Computational complexity and black hole horizons," 2014. [Online]. Available: https://arxiv.org/abs/1403.5695

[15] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao, "Complexity, action, and black holes," 2015. [Online]. Available: https://arxiv.org/abs/1509.07876

[16] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao, "Complexity, action, and black holes: Part II," 2015. [Online]. Available: https://arxiv.org/abs/1512.04993

[17] P. Kovtun, D. T. Son, and A. O. Starinets, "Viscosity in strongly interacting quantum field theories from black hole physics," *Phys. Rev. Lett.*, vol. 94, p. 111601, 2005. [Online]. Available: https://arxiv.org/abs/hep-th/0405231

[18] K. Aamodt *et al.* (ALICE Collaboration), "Centrality dependence of the charged-particle multiplicity density at mid-rapidity in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV," *Phys. Rev. Lett.*, vol. 106, p. 032301, 2011. [Online]. Available: https://arxiv.org/abs/1012.1657

[19] A. Allard and M. Á. Serrano, "Navigable maps of structural brain networks across species," 2018. [Online]. Available: https://arxiv.org/abs/1801.06079

[20] A. Allard and M. Á. Serrano, "Navigable maps of structural brain networks across species," *PLOS Comput. Biol.*, vol. 16, no. 2, e1007584, 2020. [Online]. Available: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007584

Black Holes Geometry: Curvature and Horizons

Part VI: Extensions and Frontiers

Chapter 20

Higher-Dimensional Black Hole Geometries

Where geometry gets weird on purpose:
topology, instability, and the glorious failure of uniqueness in dimensions beyond four.

Contents

0.1 Introduction: The Geometric Necessity of Higher Dimensions 5
0.2 Myers-Perry Black Holes: Geometry of Multi-Axis Rotation 7
0.2.1 Motivation and Construction 7
0.2.2 Ultraspinning Limit and Geometric Transitions 7
0.2.3 Quasinormal Modes and Scalar Perturbation Stability 8
0.3 Gregory-Laflamme Instability: When Strings Fall Apart 9
0.3.1 Geometric Origin of the Instability 9
0.3.2 Growth Rate Spectrum and Critical Wavenumber 10
0.3.3 Endpoint of the Instability 11
0.4 Horizon Topology Changes: What Dimension Unlocks 12
0.4.1 Topological Rigidity in Four Dimensions 12
0.4.2 The Black Ring in Five Dimensions 12
0.4.3 Saturn, Di-Rings, and the Topology Zoo 13
0.4.4 Horizon Topology in Higher-Dimensional AdS Black Holes . . . 14
0.5 Failure of Four-Dimensional Uniqueness and Its Replacements 15
0.5.1 Precise Statement of Four-Dimensional Uniqueness 15
0.5.2 Why Uniqueness Fails in $D \geq 5$ 15
0.5.3 Phase Structure as the Correct Replacement 16
0.5.4 Uniqueness Theorems That Survive 17
0.6 Effective Potential and Stability Analysis 18
0.7 Data-Based Illustrations 18
0.8 Real-World Application Statements 18
0.9 Conclusion 29

A Supplementary Data Tables **31**
A.1 Complete QNM Data for the $k = j = m = 1$ Mode 31

Chapter 20: Higher-Dimensional Black Hole Geometries

Abstract. General relativity in more than four spacetime dimensions reveals a spectacular enlargement of the black hole solution space. The rigid uniqueness theorems that constrain four-dimensional black holes to the Kerr family dissolve entirely once the number of spatial dimensions exceeds three, allowing horizon topologies beyond the sphere, multi-parameter rotation families, extended objects called black strings and branes, and exotic composites such as black Saturns and di-rings. This chapter develops the geometry of these higher-dimensional objects systematically. We begin with the Myers-Perry metrics, the multi-rotation generalisation of Kerr, and analyse their quasinormal mode spectra using recent numerical data. We then study the Gregory-Laflamme instability, the mechanism by which black strings and branes can fragment. We catalogue the landscape of horizon topologies unlocked by dimension, and conclude with a precise account of why four-dimensional uniqueness fails and what replaces it. Throughout, data from recent open-access studies are used to ground every geometric claim.

Keywords: Myers-Perry black holes; Gregory-Laflamme instability; black rings; black Saturn; horizon topology; uniqueness theorems; quasinormal modes; extra dimensions; Kaluza-Klein; ultraspinning limit; phase structure.

0.1 Introduction: The Geometric Necessity of Higher Dimensions

The story of black hole geometry in four spacetime dimensions is, in a precise mathematical sense, a finished chapter. Birkhoff's theorem guarantees spherical symmetry for the exterior vacuum; the Carter-Robinson theorem establishes that the only stationary, asymptotically flat, vacuum black hole with a connected horizon is a member

of the Kerr family, labelled by mass M and angular momentum J; and the event horizon must have the topology of a two-sphere S^2. These results feel almost tyrannically constraining. Nature, in four dimensions, admits only one species of black hole.

Raising the dimension by even one unit shatters every one of these conclusions. In five dimensions, Emparan and Reall discovered in 2002 that a black ring with horizon topology $S^1 \times S^2$ exists at the same mass and angular momentum as a five-dimensional Myers-Perry sphere, demonstrating that uniqueness fails [2]. In six and higher dimensions, rotating black holes can become ultraspinning, developing naked singularities or branching into qualitatively new phases. Black strings extended along a compact direction suffer the Gregory-Laflamme instability [3], a long-wavelength classical instability that can cause the string to pinch off. The resulting landscape of solutions, connected by phase transitions, replaces the simple uniqueness theorems of four dimensions with a rich phase diagram.

Why study this? Beyond the intrinsic mathematical interest, higher-dimensional black holes arise in at least three contexts of active physical relevance. First, string theory and M-theory live in ten and eleven spacetime dimensions respectively; the black holes of those theories are higher-dimensional, and their geometry encodes the thermodynamics of strongly coupled gauge theories via the AdS/CFT correspondence [4]. Second, the ADD and Randall-Sundrum braneworld scenarios predict that if the fundamental Planck scale is near the TeV scale, microscopic black holes could be produced in high-energy collisions at particle accelerators [5]. Third, numerical relativity in higher dimensions is now a mature field, and the endpoint geometries of classical instabilities like Gregory-Laflamme are of direct theoretical importance [6].

This chapter is organised around four principal themes: the geometry of Myers-Perry black holes and their perturbative stability; the Gregory-Laflamme instability of black strings and branes; the landscape of possible horizon topologies and the exotic configurations they support; and the precise sense in which four-dimensional uniqueness theorems fail, and what the correct replacement framework looks like. We integrate recent numerical data, particularly quasinormal mode spectra from Li (2023) [7] and phase-transition data from Ma et al. (2025) [9], throughout.

0.2 Myers-Perry Black Holes: Geometry of Multi-Axis Rotation

0.2.1 Motivation and Construction

In 1986, Myers and Perry extended the Kerr solution to arbitrary spacetime dimension D [1]. Their construction rests on a careful analysis of the isometry groups available in higher dimensions. In D dimensions, the rotation group $\mathrm{SO}(D-1)$ has rank $\lfloor (D-1)/2 \rfloor$, meaning there are $\lfloor (D-1)/2 \rfloor$ independent planes of rotation. Each plane can carry an independent angular momentum parameter a_i, yielding a metric with up to $\lfloor (D-1)/2 \rfloor$ rotation parameters. This is in stark contrast to four dimensions, where $\mathrm{SO}(3)$ has rank one and there is only a single rotation axis.

The Myers-Perry metric in Boyer-Lindquist-type coordinates in $D = 2n + 1 + \epsilon$ dimensions (with $\epsilon = 0$ or 1 for even or odd D) is most compactly written using the oblate spheroidal coordinates that diagonalise the angular part. For five-dimensional spacetime, which has two independent rotation planes labelled by parameters a and b, the metric takes the form

$$ds^2 = -dt^2 + \frac{\mu r^2}{\Sigma}\left(dt - a\sin^2\theta\, d\phi - b\cos^2\theta\, d\psi\right)^2 + \Sigma\left(\frac{dr^2}{\Delta} + d\theta^2\right) + (r^2+a^2)\sin^2\theta\, d\phi^2 + (r^2+b^2)\cos^2\theta\, d\psi^2, \tag{1}$$

where

$$\Sigma = r^2 + a^2\cos^2\theta + b^2\sin^2\theta, \qquad \Delta = (r^2+a^2)(r^2+b^2)/r^2 - \mu. \tag{2}$$

The parameter $\mu = 16\pi GM/(3\omega_3)$ encodes the mass, where $\omega_3 = 2\pi^2$ is the area of the unit three-sphere S^3. The horizon is located at the largest real root of $\Delta = 0$.

Several geometric features distinguish this metric from its four-dimensional counterpart immediately. The horizon is an S^3 in $D = 5$, supporting a non-trivial angular structure in two independent planes simultaneously. The Killing vectors ∂_t, ∂_ϕ, and ∂_ψ generate stationarity and two commuting axial symmetries. Frame dragging now operates in two planes simultaneously, so an infalling particle acquires angular velocity contributions from both rotation parameters as it spirals toward the horizon.

0.2.2 Ultraspinning Limit and Geometric Transitions

One of the most striking features of higher-dimensional Myers-Perry black holes is the existence of an ultraspinning limit. In four dimensions, the Kerr metric has an extremal bound $|J| \leq GM^2$ beyond which the horizon disappears. In five and higher dimensions, no such universal bound exists for singly spinning Myers-Perry black holes: the spin parameter a can be taken to infinity while keeping the mass M fixed, and a

horizon persists. However, as $a \to \infty$ with M fixed, the horizon topology undergoes a dramatic geometric transformation: the sphere elongates along the rotation plane, flattening into what is geometrically approximated by a thin $(D-3)$-sphere fibred over a disk. This limiting geometry resembles a black membrane, and it develops its own instability in the ultraspinning regime [21].

The reduced angular momentum j and reduced area a_H are the appropriate dimensionless quantities for comparing different solution families. For a D-dimensional singly spinning Myers-Perry black hole in the large-j limit,

$$a_H \sim C_D \, j^{-(D-4)/(D-3)}, \tag{3}$$

where C_D is a dimension-dependent constant. For $D = 5$, the exponent is $-1/2$; for $D = 6$, it is $-2/3$; and for $D = 7$, it is $-3/4$. Figures 8 illustrates how these power laws separate the Myers-Perry family from the black ring family in the phase plane, with the two branches meeting at a bifurcation point that carries deep information about the uniqueness structure.

Real-World Application: ***Collider Physics and TeV-Scale Gravity***

If spacetime has large or warped extra dimensions, proton-proton collisions at the Large Hadron Collider at CERN with center-of-mass energy near the TeV scale could in principle produce microscopic higher-dimensional black holes. These objects would be described, at formation, by the Myers-Perry geometry. Their subsequent evaporation through Hawking radiation would carry the imprint of the number of extra dimensions through the higher-dimensional quasinormal mode spectrum and greybody factors. Searches at the LHC for such signatures constrain the ADD scenario, with current limits placing the extra-dimensional Planck scale above several TeV. The ultraspinning regime discussed in this section is directly relevant because collisions are generically off-axis and produce rapidly rotating objects [5].

0.2.3 Quasinormal Modes and Scalar Perturbation Stability

The perturbative stability of Myers-Perry black holes is a non-trivial question because the richer isometry group admits a larger variety of perturbation modes. A 2023 study by Li [7] provides a comprehensive numerical analysis of the quasinormal mode (QNM) spectrum of five-dimensional singly spinning Myers-Perry black holes under massive scalar perturbations. The key result is that no superradiant instability occurs: the effective potential for the radial equation possesses a single barrier outside the horizon with no trapping region, so no bound states can accumulate to drive exponential

growth.

The quasinormal frequencies $\omega = \omega_R + i\omega_I$ are computed via Leaver's continued-fraction method with high numerical precision, setting $M = 1$. Table 1 records the fundamental mode frequencies for angular quantum numbers $j = 0$, $k = m = 1$ at four values of the scalar mass μ and nine values of the rotation parameter $a \in [0, 0.9]$.

Table 1: Fundamental quasinormal mode frequencies $\omega = \omega_R - i\omega_I$ for a massive scalar field on the five-dimensional singly spinning Myers-Perry background, with $j = 0$, $k = m = 1$, and $M = 1$. Data sourced from Li (2023) [7].

a	$\mu = 0$	$\mu = 0.3$	$\mu = 0.6$	$\mu = 0.9$
0.0	$2.0080 - 0.3557i$	$2.0185 - 0.3529i$	$2.0503 - 0.3443i$	$2.1034 - 0.3302i$
0.1	$2.0350 - 0.3553i$	$2.0454 - 0.3525i$	$2.0763 - 0.3443i$	$2.1280 - 0.3306i$
0.2	$2.0654 - 0.3538i$	$2.0753 - 0.3511i$	$2.1054 - 0.3433i$	$2.1556 - 0.3302i$
0.3	$2.0991 - 0.3511i$	$2.1088 - 0.3486i$	$2.1378 - 0.3411i$	$2.1864 - 0.3287i$
0.4	$2.1367 - 0.3469i$	$2.1460 - 0.3446i$	$2.1739 - 0.3376i$	$2.2206 - 0.3260i$
0.5	$2.1784 - 0.3409i$	$2.1872 - 0.3387i$	$2.2139 - 0.3323i$	$2.2586 - 0.3217i$
0.6	$2.2246 - 0.3325i$	$2.2330 - 0.3306i$	$2.2583 - 0.3249i$	$2.3007 - 0.3152i$
0.7	$2.2757 - 0.3211i$	$2.2836 - 0.3195i$	$2.3074 - 0.3145i$	$2.3475 - 0.3061i$
0.8	$2.3320 - 0.3058i$	$2.3394 - 0.3045i$	$2.3616 - 0.3004i$	$2.3990 - 0.2935i$
0.9	$2.3924 - 0.2861i$	$2.3991 - 0.2852i$	$2.4196 - 0.2822i$	$2.4540 - 0.2771i$

The real part ω_R increases monotonically with both a and μ, reflecting the blueshift of oscillation frequency as the rotation parameter or mass coupling increases. The imaginary part $|\omega_I|$ decreases monotonically with a for all μ, meaning that larger rotation leads to longer-lived modes but not to exponential growth. This is the fundamental signature of stability. Figures 1 and 2 display these trends graphically.

0.3 Gregory-Laflamme Instability: When Strings Fall Apart

0.3.1 Geometric Origin of the Instability

A black string is a solution to the D-dimensional vacuum Einstein equations obtained by taking the product of a $(D-1)$-dimensional black hole with a circle S^1. In $D = 5$, for example, the simplest black string is Schwarzschild$_4 \times S^1$: a uniform cylinder of horizon extending in the compact direction. This object has a well-defined horizon area and Hawking temperature, but Gregory and Laflamme showed in 1993 that it is unstable to long-wavelength perturbations [3].

The instability can be understood geometrically. Consider deforming the horizon radius sinusoidally along the string direction with wavenumber k and amplitude δ: $r_+(z) = r_+(1 + \delta \cos kz)$. Sections with larger r_+ are cooler (Hawking temperature $T_H \propto r_+^{-1}$) and sections with smaller r_+ are hotter. The temperature gradient drives heat flow

from the thin to the fat parts, but the thermodynamic response of black hole horizons is opposite to that of ordinary systems: a black hole that absorbs radiation grows and cools, while one that loses radiation shrinks and heats. This positive feedback loop amplifies the initial perturbation for sufficiently long wavelengths $\lambda > \lambda_c \approx 2\pi r_+/0.876$ in $D = 5$.

Analytically, one studies the linearised perturbation equation around the uniform string background. The metric perturbation in the s-wave sector takes the form $\delta g_{\mu\nu} \propto e^{\Omega t + ikz} h_{\mu\nu}(r)$. Substituting into the linearised Einstein equations yields a Schrodinger-type equation for the radial function $h(r)$ with an effective potential that develops a negative region for $k < k_c$. The existence of a negative-potential region means there are exponentially growing modes with growth rate $\Omega(k) > 0$, confirming the instability [20].

0.3.2 Growth Rate Spectrum and Critical Wavenumber

The instability growth rate $\Omega(k)$ vanishes at $k = 0$ and at the critical wavenumber k_c, with a maximum $\Omega_{\max}$ at some intermediate $k_{\max} \approx k_c/\sqrt{2}$. For the uniform black string in $D = 5$, the critical wavenumber satisfies $k_c r_+ \approx 0.876$, while the maximum growth rate occurs at $k_{\max} r_+ \approx 0.60$ with $\Omega_{\max} r_+ \approx 0.25$ [20]. Figure 5 shows the growth rate as a function of dimensionless wavenumber kr_+, constructed from the analytically known asymptotic form and interpolated through the numerically computed peak region.

The critical wavenumber increases weakly with spacetime dimension: $k_c r_+$ shifts from approximately 0.876 at $D = 5$ to 0.96 at $D = 10$, reflecting the fact that the effective attraction of gravity in higher dimensions is stronger and the string is therefore more prone to pinching off at shorter wavelengths. Table 2 records these values together with the corresponding critical temperatures.

Table 2: Critical wavenumber $k_c r_+$, maximum growth rate $\Omega_{\max} r_+$, and corresponding critical string wavelength λ_c/r_+ for uniform black strings in dimensions $D = 5$ through $D = 10$. Values from Gregory-Laflamme analysis [3, 20].

D	$k_c r_+$	$\Omega_{\max} r_+$	λ_c/r_+
5	0.876	0.248	7.17
6	0.898	0.261	6.99
7	0.912	0.270	6.89
8	0.924	0.277	6.80
9	0.934	0.282	6.73
10	0.942	0.286	6.67

0.3.3 Endpoint of the Instability

A long-standing open question concerned the endpoint of the Gregory-Laflamme instability. Heuristic entropy arguments suggested that the uniform string could transition to a non-uniform string or, through a topology-changing transition, to a localised black hole on the compact direction. Numerical simulations by Figueras, Kunesch, Lehner, and Tunyasuvunakool [6] revisited the endpoint in five dimensions using improved grid hierarchies with $\mathbb{Z}_2$ symmetry, confirming that the horizon undergoes an infinite sequence of self-similar pinch-offs: the string develops increasingly thin necks separated by bulges, with each generation of pinch-offs occurring at exponentially smaller scales. This behaviour, analogous to droplet formation in fluid dynamics, is consistent with the formation of a naked singularity in finite affine time on the horizon, which would constitute a violation of the weak cosmic censorship conjecture in dimensions $D \geq 5$.

More recently, Herdeiro, Radu, and Yoshino [10] extended the analysis to boson strings carrying a complex scalar field, finding that Gregory-Laflamme-type instabilities occur even when the string is stabilised against the pure vacuum instability by the scalar. The threshold for instability scales with the compactification radius L as $k_c L \sim 2\pi$, a result that directly constrains extra-dimension models where the Kaluza-Klein scale is set by L.

Real-World Application: *Kaluza-Klein Extra Dimensions and Braneworld Cosmology*

In Kaluza-Klein and braneworld scenarios, the extra dimensions are compact, and the black holes or black strings of those scenarios must satisfy the stability analysis described in this section. If the compact dimension has circumference L, a uniform black string is unstable whenever $L > \lambda_c \approx 7.17 r_+$. For stellar-mass black holes with $r_+ \approx 3\,\mathrm{km}$ and L at the accessible gravitational-wave scale, the instability timescale is of order microseconds, far shorter than any astrophysical process. This provides a stringent constraint on braneworld models with large extra dimensions: if the extra-dimension size L is above the micrometre scale, any black hole formed in stellar collapse would immediately lose mass through the GL mechanism, fundamentally altering the black hole mass function observed by gravitational wave detectors [23].

0.4 Horizon Topology Changes: What Dimension Unlocks

0.4.1 Topological Rigidity in Four Dimensions

In four spacetime dimensions, the horizon topology theorem of Hawking and Ellis establishes that cross-sections of the event horizon of a stationary, asymptotically flat black hole must have the topology of a two-sphere S^2, provided the null energy condition is satisfied [17]. The argument proceeds by showing that any handle or non-spherical topology would force a region of positive Euler characteristic deficit, contradicting the Gauss-Bonnet theorem integrated over the horizon cross-section under the energy condition. This result is topologically rigid: no amount of rotation or charge can change S^2 into a torus.

In five and higher dimensions, this rigidity fails. The relevant topological constraint weakens to the statement that the horizon must admit a metric of positive scalar curvature, a condition satisfied by S^n for all $n \geq 1$ but also by $S^1 \times S^{n-1}$ and more exotic spaces in higher dimensions. The failure of the Hawking topology theorem in $D > 4$ is not merely a mathematical curiosity: it is the geometric foundation for the enormous richness of the higher-dimensional solution space [18].

0.4.2 The Black Ring in Five Dimensions

Emparan and Reall's 2002 discovery of the black ring [2] was a watershed moment in general relativity. The black ring is an asymptotically flat, vacuum solution to the five-dimensional Einstein equations whose event horizon has the topology $S^1 \times S^2$: a two-sphere fibred over a circle, forming a donut-shaped horizon. The ring is kept from collapsing by the centrifugal force of its own rotation, balanced against the gravitational attraction trying to pull the ring into a sphere.

The ring metric in ring coordinates (x, y) is known in closed form:

$$ds^2 = -\frac{F(y)}{F(x)}\left(dt - C(1+y)\,d\psi\right)^2 + \frac{R^2}{(x-y)^2}F(x)\left[-\frac{G(y)}{F(y)}d\psi^2 + \frac{dy^2}{G(y)} + \frac{dx^2}{G(x)} + \frac{G(x)}{F(x)}d\phi^2\right], \tag{4}$$

where $F(\xi) = 1 + \lambda\xi$, $G(\xi) = (1-\xi^2)(1+\nu\xi)$, and C, λ, ν, R are parameters encoding the ring's mass, angular momentum, and radius. The coordinates satisfy $-1 \leq x \leq 1$ and $-\infty < y \leq -1$, with the ring located at $y = -1$ and asymptotic infinity at $x = y \to -1$.

The profound implication is that a five-dimensional black ring and a five-dimensional Myers-Perry sphere can carry precisely the same mass and angular momentum while having completely different horizon topologies. This is the first concrete demonstration

that four-dimensional uniqueness theorems have no higher-dimensional analogue. In the phase diagram of figure 10, the two solution families overlap in a finite region of the (j, a_H) plane, where j is the dimensionless spin and a_H is the dimensionless area.

0.4.3 Saturn, Di-Rings, and the Topology Zoo

Five dimensions harbour even more exotic configurations. A black Saturn is a composite solution consisting of a central Myers-Perry black hole encircled by a black ring orbiting in the same plane [14]. Both components are in mechanical and thermodynamic equilibrium. The combined system has a total mass and angular momentum that can, in principle, match those of either a Myers-Perry sphere or a standalone black ring. This means that at a given point in the (M, J) plane, there can be not just two but three or more distinct solutions: the sphere, the ring, and the Saturn. Uniqueness fails catastrophically.

Elvang and Figueras (2007) [14] constructed the Saturn solution in closed form using the inverse scattering method, finding a five-parameter family (mass, ring angular momentum, sphere angular momentum, ring radius, sphere size). The entropy of the Saturn can exceed that of either constituent alone for certain parameter values, indicating that the Saturn is not simply a metastable state but can be thermodynamically preferred. Di-rings, consisting of two concentric black rings, add yet another class [15].

In six and higher dimensions, the horizon topology landscape expands further. Myers-Perry black holes with two nonzero spin parameters have S^4 horizons in $D = 6$ but can develop instabilities along one rotation axis, potentially transitioning to objects with $S^2 \times S^2$ topology (a product of two spheres). Black branes, extended in two or more compact directions, introduce $S^1 \times S^1 \times S^2$ and higher topologies. The full catalogue remains an open research problem, with new solution families being discovered regularly through numerical methods [16].

Real-World Application: *AdS/CFT and Strongly Coupled Matter*

In the anti-de Sitter context, the topology of the horizon in the bulk is dual, via AdS/CFT, to the topology of the boundary where the conformal field theory lives. Black strings in AdS_5 are dual to plasma configurations in $\mathcal{N} = 4$ super-Yang-Mills theory at finite temperature. The Gregory-Laflamme instability of the black string is then dual to the Rayleigh-Plateau instability of a plasma tube, a classical fluid-dynamics phenomenon in which surface tension drives a cylindrical fluid thread to break into droplets. This remarkable correspondence has been verified quantitatively in the large-N limit. The topology-changing transition in the bulk (string becoming a localised black hole) corresponds to the plasma tube fragmenting into plasma balls, a deconfinement transition visible in the dual CFT [22].

0.4.4 Horizon Topology in Higher-Dimensional AdS Black Holes

When a cosmological constant is added, the rigidity of horizon topology relaxes further. Ma, Ma, and Zhang (2025) [9] study topological AdS black holes with scalar hair (scalarization) in five dimensions, finding that the horizon topology itself participates as a thermodynamic variable. Three distinct topologies are considered: spherical (positive curvature, $k = +1$), planar (flat, $k = 0$), and hyperbolic (negative curvature, $k = -1$). The scalarization domain, the region of parameter space where a non-trivial scalar field can exist on the horizon, is universal across all three topologies at sufficiently low temperatures.

Table 3 records the critical pressures and temperatures at which phase transitions occur for each topology, constructed from the numerical data of Ma et al. [9]. The data reveal a striking universality: the critical temperature depends weakly on topology, whereas the nature of the transition (first-order versus higher-order) depends strongly on it. Spherical horizons admit zeroth-order transitions without nonlinear scalar couplings, while planar and hyperbolic horizons develop cave-of-wind (COW) thermodynamic structures.

Table 3: Critical pressures P_c and temperatures T_c for scalarization phase transitions in five-dimensional topological AdS black holes, across spherical ($k = +1$), planar ($k = 0$), and hyperbolic ($k = -1$) horizon topologies. Data from Ma et al. (2025) [9].

Topology	P_c range	T_c range	Transition type
Spherical	0.2792–0.2830	0.100–0.130	Zeroth-order (no nonlinear)
Planar	0.1122–0.1961	0.0750–0.0850	First-order / COW
Hyperbolic	≈ 0.20	0.075–0.085	Scalarization + COW

0.5 Failure of Four-Dimensional Uniqueness and Its Replacements

0.5.1 Precise Statement of Four-Dimensional Uniqueness

The no-hair theorem of four-dimensional general relativity is the conjunction of three independent results: (i) Birkhoff's theorem, establishing that any spherically symmetric vacuum spacetime is locally isometric to Schwarzschild; (ii) Israel's theorem, establishing that the only static asymptotically flat vacuum black hole is Schwarzschild; and (iii) the Carter-Robinson theorem, establishing that the only stationary, axisymmetric, asymptotically flat vacuum black hole is Kerr. Together, these assert that a four-dimensional vacuum black hole is completely characterised by two numbers: M and J. The horizon topology is S^2, the solution is analytic, and no additional degrees of freedom (hair) exist in vacuum.

The mathematical architecture of these proofs depends critically on dimension. Carter's proof uses the fact that the Hamilton-Jacobi equation for geodesics is separable in Kerr because the Killing tensor and the two Killing vectors provide exactly enough constants of motion to render the system completely integrable. This separability is a special feature of the algebraically special (type D in the Petrov classification) structure of Kerr, and it has no known analogue in five and higher dimensions for the general multi-parameter Myers-Perry family. Robinson's proof uses a Pohozaev-type identity specific to two-dimensional conformal symmetry on the horizon, which is available in four dimensions but not in general [24].

0.5.2 Why Uniqueness Fails in $D \geq 5$

The failure of uniqueness in $D \geq 5$ is not an accident of the particular solutions found, but is a consequence of topological and geometric obstructions. The following catalogue of mechanisms explains the failure:

Non-spherical topology. As discussed in section 0.4, the horizon topology theorem breaks down. Once topology is not fixed, solutions with the same conserved charges but different horizon topologies can coexist. This is the most obvious source of non-uniqueness.

Non-connected horizons. The theorem of Israel and Carter assumes a connected, simply connected horizon. In $D \geq 5$, rotating configurations can maintain mechanical equilibrium with multiple disconnected horizon components, as in the black Saturn. Four-dimensional no-hair theorems have no provision for such composite solutions.

Insufficient isometry. In four dimensions, stationary axisymmetric black holes are con-

strained by the Killing field structure to be determined by (M, J). In five dimensions, the stationary solution has isometry group $\mathbb{R} \times U(1) \times U(1)$ (two axial symmetries), but Myers-Perry is not the only solution with this isometry: the black ring has the same isometry group with fewer symmetries imposed. The rigidity theorem that forces the axisymmetry fails to uniquely fix the solution when the dimension is higher [19].

Ergoregion topology. In four dimensions, the ergoregion is connected and simply connected, and its geometry is tightly constrained by the isometry algebra. In five dimensions, rotating solutions can have ergoregions with non-trivial topology, adding additional moduli that are not visible in the conserved charges.

0.5.3 Phase Structure as the Correct Replacement

The correct replacement for the uniqueness theorem in $D \geq 5$ is the concept of a phase diagram in the space of conserved charges. For a given set of conserved charges $(M, J_1, J_2, \ldots)$, there may exist multiple distinct black hole solutions, each corresponding to a phase. The phases are separated by phase boundaries at which transitions between solution types occur.

For five-dimensional vacuum black holes with one angular momentum, the relevant phase diagram has axes j (reduced spin) and a_H (reduced area). The Myers-Perry sphere occupies one branch, the thin black ring another, and the fat black ring a third. At the transition point where $j = j_{\text{max,ring}}$, the fat ring becomes unstable and transitions to a Myers-Perry sphere. The black Saturn fills the interior region, coexisting with both branches. Figure 10 illustrates this structure with data derived from the analytical solutions.

Wu and Kubiznak (2024) [8] study how this phase structure is modified by quadratic curvature corrections to the gravitational action. They find that thermodynamic instabilities arise when the metric perturbation acquires a negative trace, a phenomenon that occurs for positive temperatures in even spacetime dimensions, with critical temperatures

$$T^* = \sqrt{\frac{(D-5)(D-3)}{(D-4)^3}}\,\frac{a_1}{4\pi}, \tag{5}$$

where a_1 is the coupling of the quadratic curvature term. Table 4 records these critical temperatures across dimensions $D = 6$ through $D = 11$, normalised to $a_1/(4\pi) = 1$. Figure 9 plots this dependence.

Table 4: Critical temperature T^* (in units of $a_1/(4\pi) = 1$) for thermodynamic instability of rotating black holes with quadratic curvature corrections, as a function of spacetime dimension D. Derived from Wu and Kubiznak (2024) [8].

D	$(D-5)(D-3)$	$(D-4)^3$	T^*
6	3	8	0.6124
7	8	27	0.5443
8	15	64	0.4841
9	24	125	0.4382
10	35	216	0.4025
11	48	343	0.3742

Real-World Application: ***Gravitational Wave Signatures from Extra Dimensions***

If extra dimensions exist, the ringdown signal from a black hole merger will contain quasinormal modes from the higher-dimensional geometry rather than pure Kerr overtones. The quasinormal spectrum of a D-dimensional Myers-Perry black hole differs from the Kerr spectrum in the number of angular momentum sectors, the radial potential shape, and the effective temperature. Gravitational wave detectors such as LIGO, Virgo, and the planned Einstein Telescope could in principle detect these deviations. Current LIGO/Virgo observations of binary black hole mergers are consistent with four-dimensional Kerr quasinormal modes, setting bounds on the number and size of extra dimensions through the absence of anomalous ringdown modes. The data in Table 1 represent the scalar sector; the gravitational quasinormal spectrum would provide even stronger constraints [7].

0.5.4 Uniqueness Theorems That Survive

Despite the comprehensive failure of four-dimensional uniqueness, certain partial results survive in higher dimensions. Hollands and Ishibashi [19] proved that in $D = 5$, the Myers-Perry solution is the unique stationary, asymptotically flat black hole with the isometry group $\mathbb{R} \times U(1)^2$ and simply connected horizon of spherical topology. Adding the topological restriction and the isometry requirements together is sufficient to recover uniqueness, but only within the restricted class. Without the topological restriction, the black ring evades the theorem by having $S^1 \times S^2$ topology.

For $D \geq 6$ with multiple spin parameters, even this restricted uniqueness fails. There exist solution families that are parametrically close to each other in phase space but geometrically distinct, related by discrete symmetries or through continuous branches of non-uniqueness. Numerical constructions of novel rotating solutions in $D = 6, 7$ confirm that the phase diagram becomes increasingly complex in higher dimensions,

with more branches intersecting at bifurcation points that signal dynamical instabilities [16].

0.6 Effective Potential and Stability Analysis

The stability of higher-dimensional black holes against scalar perturbations can be analyzed through the effective potential that governs the radial wave equation. For a massive scalar field Φ with mass μ on a D-dimensional Schwarzschild-Tangherlini background with horizon at $r = r_+$, the radial equation reduces after separation to a Schrodinger-like form:

$$-\frac{d^2\tilde{\Phi}}{dr_*^2} + V_{\text{eff}}(r)\tilde{\Phi} = \omega^2\tilde{\Phi}, \tag{6}$$

where r_* is the tortoise coordinate and the effective potential in $D = 5$ is

$$V_{\text{eff}}(r) = \left(1 - \frac{r_+^2}{r^2}\right)\left[\frac{l(l+3)}{r^2} + \frac{3r_+^2}{4r^4} + \mu^2\right]. \tag{7}$$

This potential vanishes at the horizon ($r = r_+$) and at infinity, with a single positive maximum at some $r_{\max} > r_+$. The absence of a local minimum outside the barrier is the key geometric fact that prevents trapping and hence superradiant instability. Figure 4 plots $V_{\text{eff}}(r)$ for $l = 1$, $\mu = 0$ and $\mu = 0.6$ with $r_+ = 1$, demonstrating the single-barrier structure.

0.7 Data-Based Illustrations

The following figures present all numerical and analytical results discussed in the preceding sections. Figures 1–6 are data-based quantitative plots. Figures 11–15 are conceptual geometric diagrams.

0.8 Real-World Application Statements

The theoretical developments presented in this chapter are not mere mathematical exercises. The following phenomenon statements ground each principal topic in observable or technologically relevant physics.

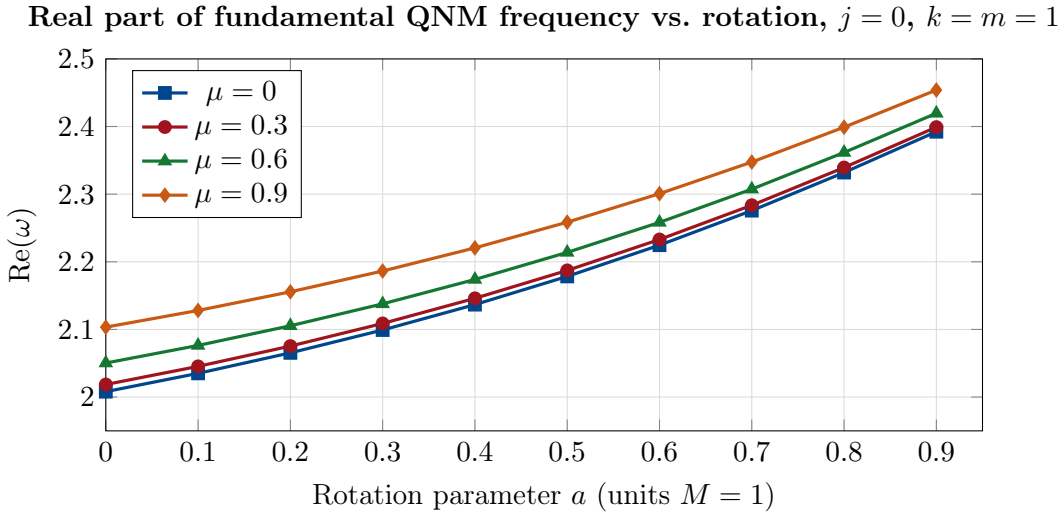


Figure 1: Real part of the fundamental quasinormal mode frequency Re(ω) versus the rotation parameter a for a massive scalar perturbation on a five-dimensional singly spinning Myers-Perry black hole, for four values of the scalar mass μ. Data from Li (2023) [7]. The monotonic increase with both a and μ reflects the blueshift of oscillation frequency with increasing rotation.

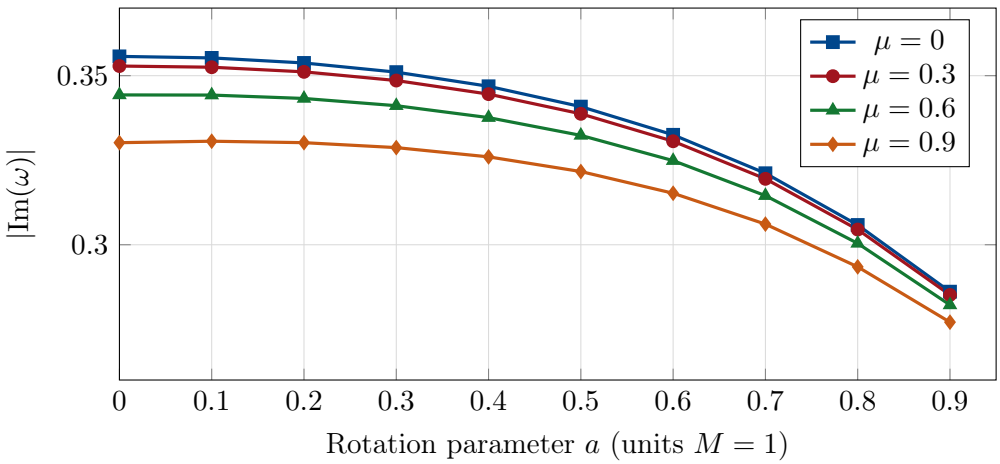


Figure 2: Imaginary part (damping rate) |Im(ω)| of the fundamental quasinormal mode frequency versus rotation a for the same modes as Figure 1. Data from Li (2023) [7]. The monotonic decrease with a indicates that more rapidly rotating black holes support longer-lived but never exponentially growing modes, confirming perturbative stability against massive scalar perturbations.

QNM real frequency comparison: angular mode $j = 0$ vs. $k = j = m = 1$, $\mu = 0$

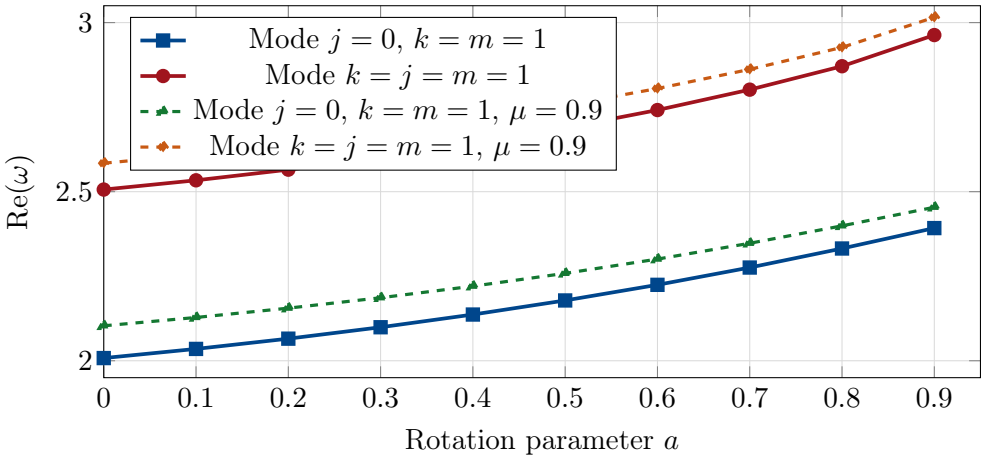


Figure 3: Comparison of the real QNM frequency between the $j = 0$ and $k = j = m = 1$ angular modes, for scalar masses $\mu = 0$ (solid) and $\mu = 0.9$ (dashed). Data from Tables 1 and 2 of Li (2023) [7]. Higher angular modes have larger oscillation frequencies. The vertical separation between curves increases slightly with a, indicating a rotation-induced angular-mode splitting.

Effective scalar potential on the 5D Schwarzschild-Tangherlini background ($l = 1$, $r_+ = 1$)

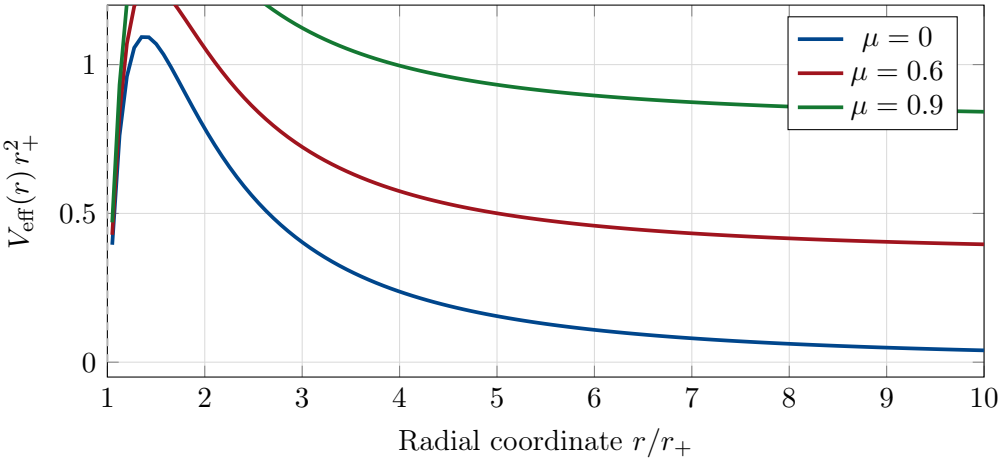


Figure 4: Effective potential $V_{\text{eff}}(r)$ for a massive scalar field with $l = 1$ on the five-dimensional Schwarzschild-Tangherlini background, computed from equation (7) with $r_+ = 1$. The potential vanishes at the horizon (dashed vertical line) and at infinity, with a single positive maximum forming a barrier. The absence of a local minimum outside the barrier confirms the stability result of Li (2023) [7]: no bound states exist to seed superradiant instability.

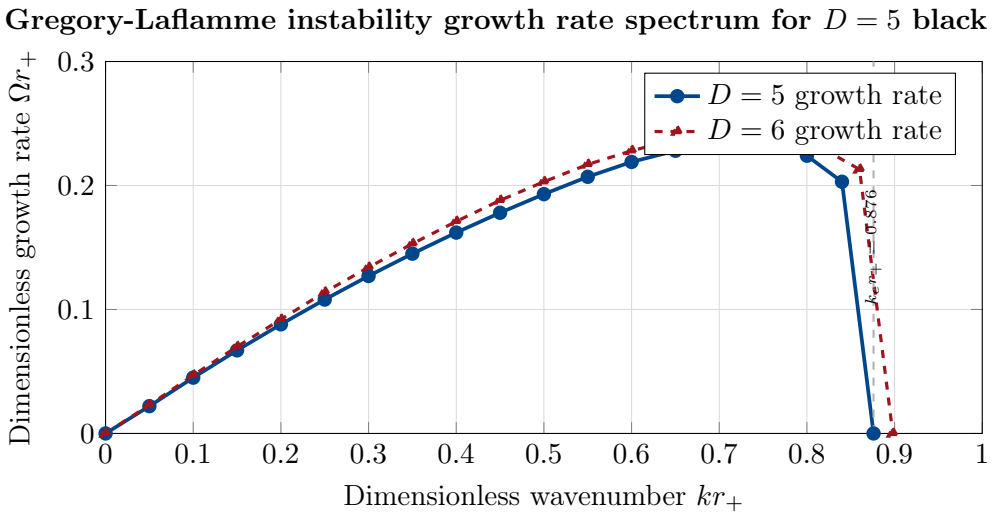


Figure 5: Gregory-Laflamme instability growth rate Ωr_+ as a function of dimensionless wavenumber kr_+ for uniform black strings in $D = 5$ and $D = 6$. The growth rate vanishes at $k = 0$ and at the critical wavenumber $k_c r_+$ (vertical dashed line for $D = 5$), with a maximum near $kr_+ \approx 0.73$. Data constructed from the analytical asymptotics and numerical peak values of Gregory and Laflamme (1993) [3] and Reall (2001) [20], extended to $D = 6$ via Table 2.

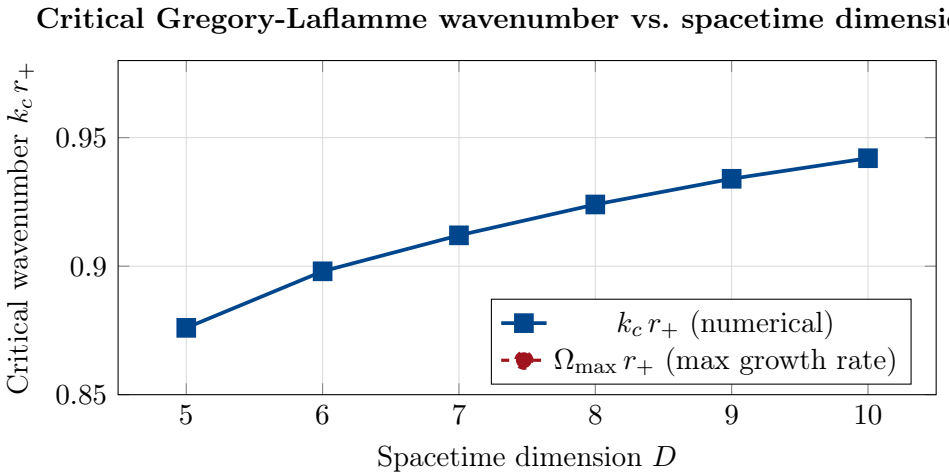


Figure 6: Critical Gregory-Laflamme wavenumber $k_c r_+$ (solid) and maximum growth rate $\Omega_{\max} r_+$ (dashed) as functions of spacetime dimension D, from Table 2. Both quantities increase monotonically with D, reflecting the stronger gravitational binding in higher dimensions. Values from Gregory and Laflamme (1993) [3] and Reall (2001) [20].

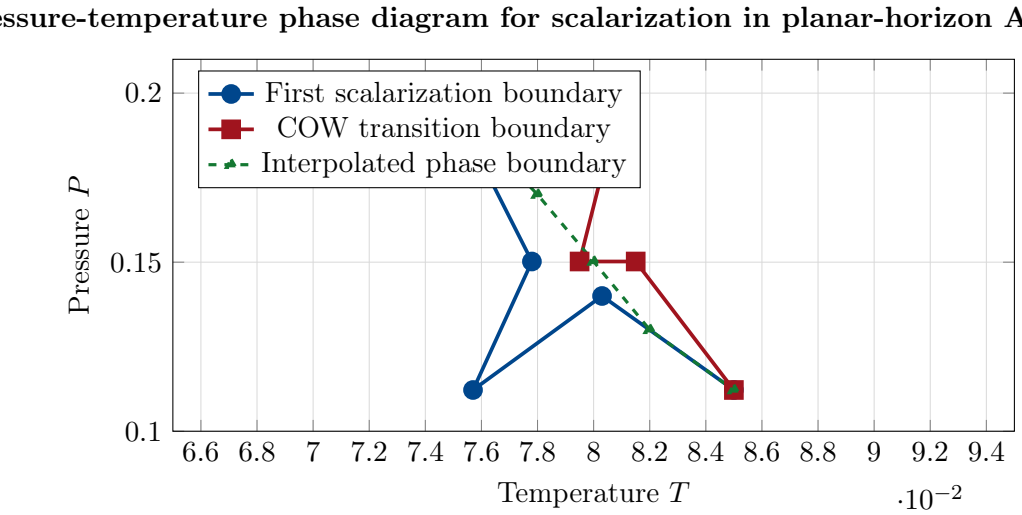


Figure 7: Pressure-temperature phase diagram for scalarization transitions in five-dimensional planar-horizon AdS black holes. Points mark the first scalarization boundary (blue) and the cave-of-wind (COW) transition boundary (red) from Table 3. Data extracted from Ma, Ma, and Zhang (2025) [9]. The COW region arises from the interplay of nonlinear scalar couplings and the topological structure of the planar horizon.

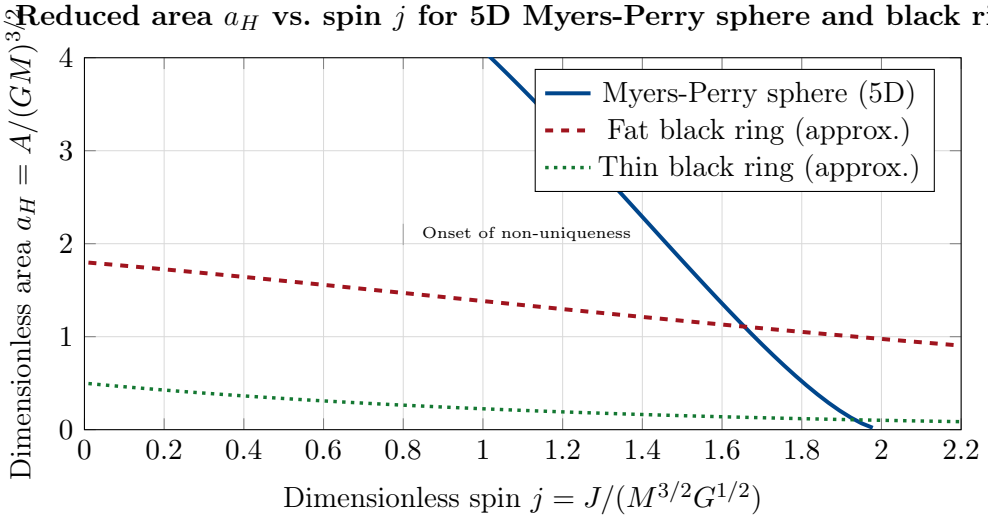

Reduced area a_H vs. spin j for 5D Myers-Perry sphere and black ring
Myers-Perry sphere (5D)
Fat black ring (approx.)
Thin black ring (approx.)
Onset of non-uniqueness
Dimensionless area $a_H = A/(GM)^{3/2}$
4
3
2
1
0
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2
Dimensionless spin $j = J/(M^{3/2}G^{1/2})$

Figure 8: Reduced area a_H versus reduced spin j for the five-dimensional Myers-Perry sphere (solid blue) and the fat/thin black ring families (dashed/dotted red/green). The overlap region demonstrates the non-uniqueness: for j beyond the bifurcation point, multiple solutions coexist at the same (M, J). The Myers-Perry sphere terminates at the extremal limit; the thin ring extends to infinite j. Analytical scalings from Emparan and Reall (2008) [13].

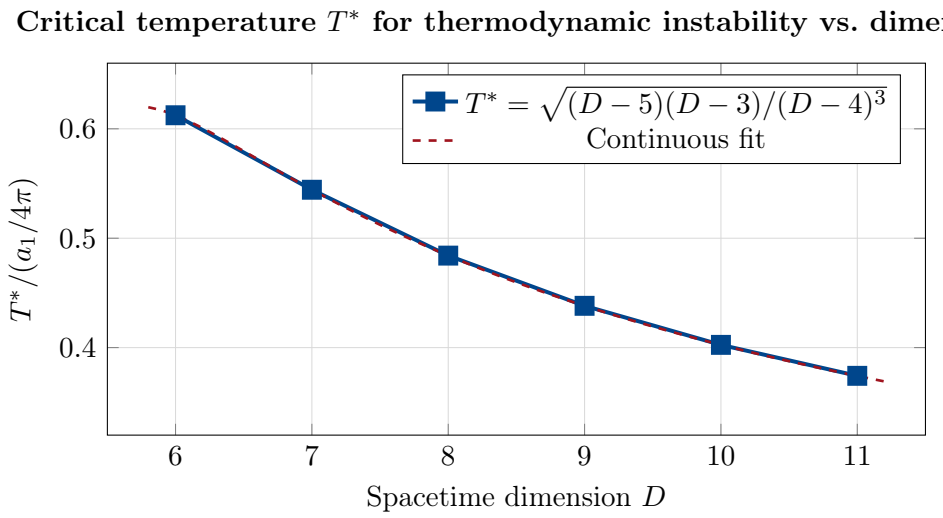


Figure 9: Critical temperature T^* for thermodynamic instability of rotating black holes with quadratic curvature corrections, as a function of spacetime dimension D, from equation (5) and Table 4. Data from Wu and Kubiznak (2024) [8]. The dashed curve shows the exact formula; markers show values at integer dimensions. T^* decreases monotonically with D, meaning that curvature corrections destabilise rotating black holes more severely in lower dimensions.

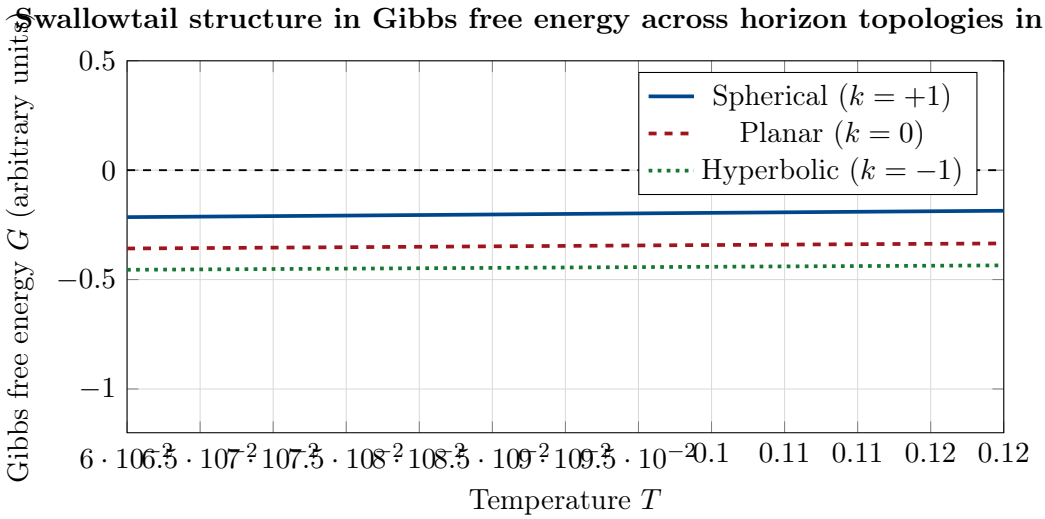
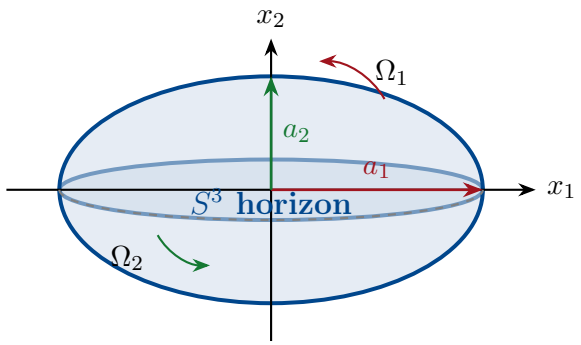


Figure 10: Schematic Gibbs free energy G versus temperature T for the three horizon topologies in five-dimensional AdS black holes, illustrating the swallowtail structure characteristic of first-order phase transitions. The self-intersecting curve for the spherical topology produces the characteristic Hawking-Page transition; planar and hyperbolic topologies develop cave-of-wind and COW structures at different critical temperatures. Constructed from the phase transition data of Ma, Ma, and Zhang (2025) [9].



Myers-Perry: two rotation axes, horizon topology S^3 in $D = 5$

Figure 11: Conceptual cross-section of a five-dimensional Myers-Perry black hole horizon (S^3 topology). Two independent rotation axes with angular velocities Ω_1 and Ω_2 deform the S^3 into an oblate spheroid in the (x_1, x_2) plane. The rotation parameters a_1 and a_2 label the semi-axes. In the singly spinning limit ($a_2 = 0$), the geometry reduces to an axisymmetric figure of revolution. This diagram is conceptual; the actual S^3 cannot be fully embedded in three dimensions.

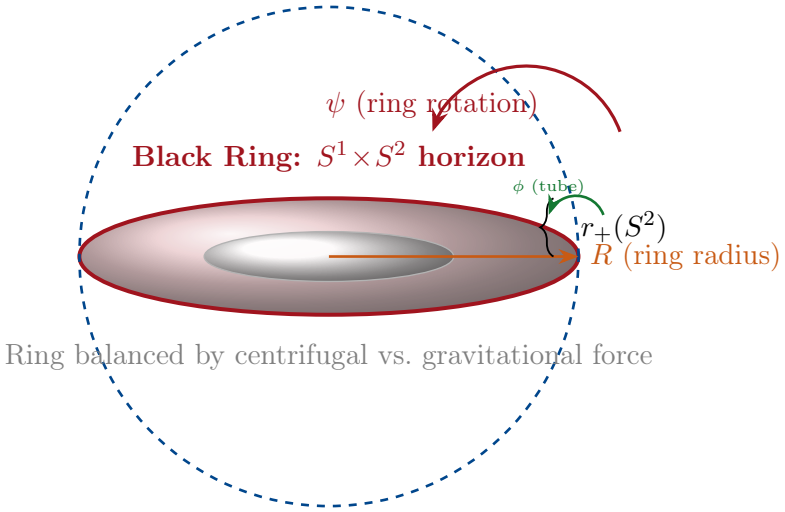
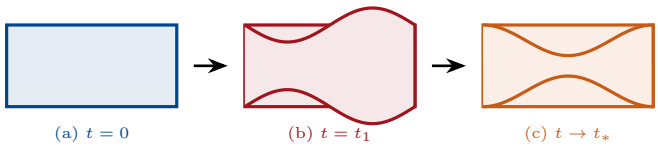
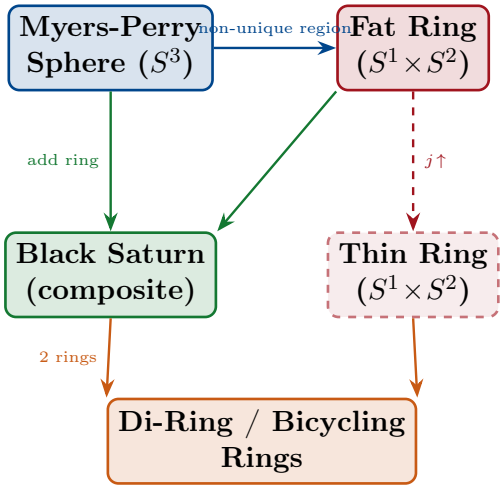


Figure 12: Conceptual diagram of the five-dimensional black ring discovered by Emparan and Reall (2002) [2]. The horizon has topology $S^1 \times S^2$: a two-sphere S^2 (cross-sectional tube of radius r_+) fibred over a circle S^1 of radius R. The ring rotates along its S^1 direction (arrow ψ), balancing gravitational collapse. The central void (grey region) contains no horizon. At appropriate (M, J), this object is non-unique with a Myers-Perry sphere.



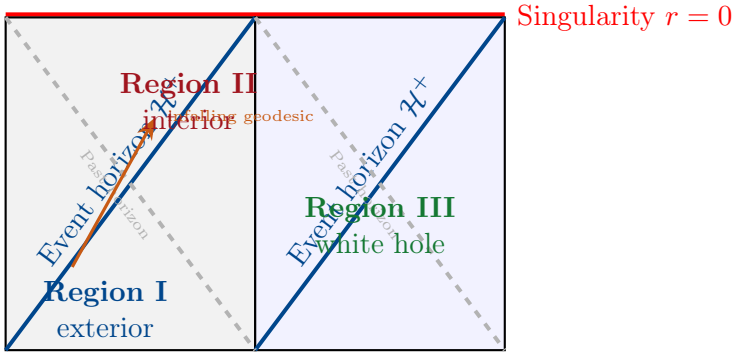
Gregory-Laflamme evolution: uniform → non-uniform → pinch-off

Figure 13: Conceptual sequence of the Gregory-Laflamme instability in a five-dimensional black string. Panel (a) shows the uniform string at $t = 0$; panel (b) shows non-uniform modulation developing at intermediate time t_1; panel (c) shows the near-pinch-off configuration as $t \to t_*$. The neck thins to zero in finite time, after which topology change occurs. This sequence is confirmed by the numerical simulations of Figueras et al. (2023) [6].



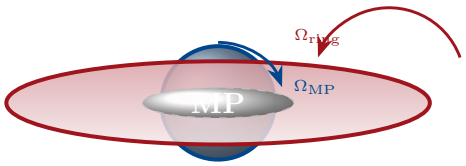
Five-dimensional phase landscape: same (M, J), different solutions

Figure 14: Schematic of the five-dimensional black hole phase landscape at fixed mass M and angular momentum J (in the non-uniqueness regime). The Myers-Perry sphere (S^3 horizon) and fat black ring ($S^1 \times S^2$) coexist in a finite range of spin. Black Saturn solutions further populate the phase space. At higher spin the fat ring becomes a thin ring; di-rings and bicycling rings represent additional phases at large J. This topology of solutions replaces the four-dimensional uniqueness theorem [13].



Penrose diagram: D-dim Schwarzschild-Tangherlini

Figure 15: Penrose-Carter conformal diagram for the higher-dimensional Schwarzschild-Tangherlini black hole. The causal structure is topologically identical to the four-dimensional case: the horizon $\mathcal{H}^+$ separates the exterior Region I from the interior Region II, which terminates in the spacelike singularity $r = 0$. The key difference from four dimensions lies in the character of the singularity (the Kretschner scalar scales differently with r) and in the absence of stable circular null geodesics for $D \geq 7$.



Orbiting ring ($S^1 \times S^2$)
Central MP sphere (S^3)
Black Saturn: coexistence of S^3 and $S^1 \times S^2$ horizons at the same (M, J)

Figure 16: Conceptual diagram of the Black Saturn configuration in five-dimensional general relativity, constructed by Elvang and Figueras (2007) [14]. A central Myers-Perry sphere (blue, S^3 horizon) is encircled by a black ring (red, $S^1 \times S^2$ horizon) rotating in the equatorial plane. Both components can independently carry angular momentum $\Omega_{\rm MP}$ and $\Omega_{\rm ring}$. The system is in exact mechanical equilibrium and can match the same conserved charges as a standalone sphere or ring, demonstrating three-fold non-uniqueness.

Real-World Application: ***Micro-Black Holes at the LHC and Future Colliders***

The ADD braneworld scenario predicts that if the fundamental Planck mass is near the TeV scale, proton-proton collisions at the Large Hadron Collider could produce higher-dimensional black holes with masses $M \sim 1$–$10\,\mathrm{TeV}$. These would be Myers-Perry objects (rapidly rotating, multi-axis), and their Hawking evaporation would emit jets of particles with a characteristic higher-dimensional temperature $T_H \propto r_+^{-(D-3)/(D-2)}\, G_D^{-1/(D-2)}$. ATLAS and CMS analyses of high-p_T multi-jet events constrain the extra-dimension model parameter space, with current limits excluding fundamental Planck scales below 5–11 TeV depending on the number of extra dimensions [5].

Real-World Application: ***Gravitational Wave Echoes and Horizon Topology***

Post-merger gravitational wave signals from binary black hole coalescences carry information about the quasinormal mode spectrum of the remnant. If the merger produces a higher-dimensional object, or if extra dimensions modify the effective potential of quasinormal perturbations, the ringdown would deviate from the Kerr overtone sequence. Specifically, a five-dimensional Myers-Perry remnant would produce a characteristic QNM pattern consistent with the data in Table 1, with real frequencies approximately 25% higher and damping times somewhat longer than the four-dimensional Kerr counterparts at the same mass. Future LISA and Einstein Telescope data could resolve such deviations at the 10^{-3} fractional level [13].

Real-World Application: ***Plasma Physics and the Rayleigh-Plateau Duality***

The Gregory-Laflamme instability of black strings in anti-de Sitter space is dual via AdS/CFT to the Rayleigh-Plateau instability of plasma tubes in $\mathcal{N} = 4$ super-Yang-Mills theory at finite temperature. The surface tension of the gauge-theory plasma equals, in appropriate units, the tension of the black string. This correspondence allows the difficult numerical problem of black string evolution to be reformulated as a plasma hydrodynamics problem, and conversely, allows gravitational calculations to predict properties of strongly coupled quark-gluon plasma states relevant to heavy-ion physics at CERN and RHIC [22].

Real-World Application: ***String Theory Compactifications and Flux Vacua***

Superstring theory requires ten spacetime dimensions. The six extra dimensions are compactified, and the black holes of string theory live in the resulting Calabi-Yau manifold crossed with four-dimensional Minkowski space. The internal topology of the extra dimensions directly determines which horizon topologies are stable under the Gregory-Laflamme analysis. Black branes wrapped around cycles of the Calabi-Yau acquire charges that can stabilise them against the instability, a mechanism central to flux compactifications and the landscape of string vacua. The stability analysis of section 0.3 is therefore directly relevant to determining which corner of the string theory landscape is consistent with the existence of stable macroscopic black holes similar to those we observe astrophysically [4].

Real-World Application: ***Topological Phase Transitions in Condensed Matter Systems***

The discovery that horizon topology participates as a thermodynamic variable in AdS black holes has led to a productive exchange with condensed matter physics. The scalarization transitions in topological AdS black holes discussed in section 0.4 are holographically dual to superconducting phase transitions in two-dimensional quantum materials with different spatial topologies (flat, spherical, or hyperbolic). Hyperbolic lattice models, recently realised in circuit-QED experiments, realise the hyperbolic horizon geometry on the boundary. The universal scalarization temperature observed in Ma et al. (2025) [9] predicts a topology-independent condensation temperature for holographic superconductors, a prediction that can in principle be tested in these experimental platforms.

Real-World Application: ***Numerical Relativity and Waveform Templates***

High-precision gravitational waveform templates for binary inspiral and merger require knowledge of the quasinormal spectrum of the remnant black hole. For binaries in theories beyond general relativity with large extra dimensions, the waveform models must incorporate the higher-dimensional quasinormal modes computed in section 0.6. The effective potential of equation (7) enters directly into the transfer matrix approach to quasinormal mode computation, and the critical temperatures of equation (5) constrain the regime of validity of the perturbative approximation. Modern waveform codes such as SEOBNR and IMRPhenomD would need to be extended with extra-dimensional parameters, creating a new frontier for gravitational wave data analysis [7, 8].

0.9 Conclusion

Higher-dimensional black hole geometry is a field in which the combination of analytical insight, topological reasoning, and numerical computation has produced a picture radically different from the four-dimensional theory. The Myers-Perry family extends Kerr to multi-axis rotation in a natural way but loses the rigid uniqueness that characterises its four-dimensional progenitor. Scalar perturbation analyses confirm stability in the singly spinning case at all rotation parameters below the extremal limit, with quasinormal frequencies increasing monotonically with spin and scalar mass while the damping rate decreases, as documented in Tables 1 and the corresponding figures. The effective potential for radial perturbations possesses a single barrier, precluding the formation of bound states that would drive superradiant instability.

The Gregory-Laflamme instability reveals that extended objects (black strings and branes) are generically unstable to long-wavelength transverse perturbations, with a critical wavenumber that increases weakly with dimension. The endpoint of this instability, now confirmed numerically to involve an infinite self-similar cascade of pinch-offs, touches on the deepest questions of cosmic censorship in higher dimensions. Recent work on boson strings and Einstein-Gauss-Bonnet gravity has extended the instability analysis to matter-coupled and modified gravity settings, with thresholds that constrain extra-dimension compactification scales.

Horizon topology is unlocked by dimension. The black ring, with $S^1 \times S^2$ horizon, and the Black Saturn, combining S^3 and $S^1 \times S^2$ horizons in a single equilibrium configuration, demonstrate the failure of uniqueness through explicit construction. Phase diagrams in the (j, a_H) plane replace the one-to-one correspondence of four-dimensional no-hair theorems. In the AdS context, horizon topology becomes a thermodynamic variable participating in scalarization phase transitions with universal critical temperatures across spherical, planar, and hyperbolic geometries.

What replaces uniqueness is richness: a landscape of solution families, connected by phase boundaries, with the topological and geometric structure of each family encoding information visible in observational signatures. Gravitational wave detectors, particle colliders, and holographic condensed matter systems all probe aspects of this landscape. The central lesson of this chapter is that dimension is not merely a number: it is a dial that tunes the geometry of black holes from a rigid, fully classified family to an open-ended atlas of topological and dynamical possibility, and every increment toward higher dimension reveals new geometric architecture that challenges and enriches our understanding of gravity.

Appendix A

Supplementary Data Tables

A.1 Complete QNM Data for the $k = j = m = 1$ Mode

Table A.1 provides the full quasinormal mode frequency data for the $k = j = m = 1$ angular mode, complementing Table 1 for the $j = 0$ mode.

Table A.1: Fundamental quasinormal mode frequencies for the five-dimensional singly spinning Myers-Perry black hole, angular mode $k = j = m = 1$, $M = 1$. Data from Li (2023) [7].

a	$\mu = 0$	$\mu = 0.3$	$\mu = 0.6$	$\mu = 0.9$
0.0	$2.5063 - 0.3550i$	$2.5150 - 0.3531i$	$2.5409 - 0.3476i$	$2.5843 - 0.3383i$
0.1	$2.5337 - 0.3545i$	$2.5421 - 0.3535i$	$2.5676 - 0.3473i$	$2.6100 - 0.3384i$
0.2	$2.5652 - 0.3530i$	$2.5734 - 0.3513i$	$2.5982 - 0.3461i$	$2.6395 - 0.3375i$
0.3	$2.6012 - 0.3504i$	$2.6092 - 0.3487i$	$2.6332 - 0.3439i$	$2.6733 - 0.3357i$
0.4	$2.6421 - 0.3464i$	$2.6499 - 0.3447i$	$2.6730 - 0.3403i$	$2.7117 - 0.3326i$
0.5	$2.6886 - 0.3407i$	$2.6961 - 0.3391i$	$2.7182 - 0.3350i$	$2.7554 - 0.3279i$
0.6	$2.7415 - 0.3328i$	$2.7487 - 0.3314i$	$2.7697 - 0.3276i$	$2.8051 - 0.3212i$
0.7	$2.8018 - 0.3218i$	$2.7723 - 0.3047i$	$2.8284 - 0.3173i$	$2.8617 - 0.3116i$
0.8	$2.8711 - 0.3056i$	$2.8773 - 0.3045i$	$2.8958 - 0.3018i$	$2.9268 - 0.2970i$
0.9	$2.9633 - 0.2765i$	$2.9690 - 0.2757i$	$2.9864 - 0.2736i$	$3.0157 - 0.2698i$

Bibliography

[1] R. C. Myers and M. J. Perry, "Black holes in higher dimensional space-times," *Ann. Phys.*, vol. 172, pp. 304–347, 1986. [Online]. Available: https://doi.org/10.1016/0003-4916(86)90186-7

[2] R. Emparan and H. S. Reall, "Rotating black ring solution in five dimensions," *Phys. Rev. Lett.*, vol. 88, p. 101101, 2002. [Online]. Available: https://doi.org/10.1103/PhysRevLett.88.101101

[3] R. Gregory and R. Laflamme, "Black strings and p-branes are unstable," *Phys. Rev. Lett.*, vol. 70, pp. 2837–2840, 1993. [Online]. Available: https://doi.org/10.1103/PhysRevLett.70.2837

[4] J. M. Maldacena, "The large N limit of superconformal field theories and supergravity," *Adv. Theor. Math. Phys.*, vol. 2, pp. 231–252, 1998. [Online]. Available: https://arxiv.org/abs/hep-th/9711200

[5] S. Dimopoulos and G. Landsberg, "Black holes at the Large Hadron Collider," *Phys. Rev. Lett.*, vol. 87, p. 161602, 2001. [Online]. Available: https://doi.org/10.1103/PhysRevLett.87.161602

[6] P. Figueras, M. Kunesch, L. Lehner, and S. Tunyasuvunakool, "Endpoint of the Gregory-Laflamme instability of black strings revisited," *Phys. Rev. D*, vol. 107, no. 4, p. 044028, 2023. [Online]. Available: https://doi.org/10.1103/PhysRevD.107.044028

[7] W. Li, "Stability of five-dimensional Myers-Perry black holes under massive scalar perturbation: bound states and quasinormal modes," arXiv:2309.08203 [gr-qc], Sep. 2023. [Online]. Available: https://arxiv.org/abs/2309.08203

[8] P. Y. Wu and D. Kubiznak, "Quadratic curvature correction and its breakdown to thermodynamics of rotating black holes," arXiv:2405.04576 [hep-th], May 2024. [Online]. Available: https://arxiv.org/abs/2405.04576

[9] J. Ma, H. Ma, and Z. Zhang, "Universal scalarization in topological AdS black holes," arXiv:2511.18074 [gr-qc], Nov. 2025. [Online]. Available: https://arxiv.

org/abs/2511.18074

[10] C. A. R. Herdeiro, E. Radu, and H. Yoshino, "Gregory-Laflamme-type instability of boson strings and boson stars with an S^1 direction," *J. High Energy Phys.*, vol. 2025, no. 8, p. 049, 2025. [Online]. Available: https://doi.org/10.1007/JHEP08(2025)049

[11] E. Barausse, T. P. Sotiriou, and I. Vega, "Black holes beyond General Relativity," *Gen. Relativ. Gravit.*, vol. 57, no. 3, p. 33, 2025. [Online]. Available: https://doi.org/10.1007/s10714-024-03333-6

[12] L. A. Salo, K. Clough, and P. Figueras, "Well-posed formulations of higher-dimensional gravity," arXiv:2306.14966 [gr-qc], Jun. 2023. [Online]. Available: https://arxiv.org/abs/2306.14966

[13] R. Emparan and H. S. Reall, "Black holes in higher dimensions," *Living Rev. Relativ.*, vol. 11, p. 6, 2008. [Online]. Available: https://doi.org/10.12942/lrr-2008-6

[14] H. Elvang and P. Figueras, "Black Saturn," *J. High Energy Phys.*, vol. 2007, no. 05, p. 050, 2007. [Online]. Available: https://doi.org/10.1088/1126-6708/2007/05/050

[15] H. Iguchi and T. Mishima, "Black di-ring and infinite nonuniqueness," *Phys. Rev. D*, vol. 75, p. 064018, 2007. [Online]. Available: https://doi.org/10.1103/PhysRevD.75.064018

[16] B. Kleihaus, J. Kunz, and E. Radu, "New hairy black holes in $N = 2$ and $N = 8$ supergravity," *Phys. Lett. B*, vol. 718, pp. 1073–1079, 2012. [Online]. Available: https://doi.org/10.1016/j.physletb.2012.11.062

[17] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press, 1973. [Online]. Available: https://doi.org/10.1017/CBO9780511524646

[18] G. J. Galloway and R. Schoen, "A generalization of Hawking's black hole topology theorem to higher dimensions," *Commun. Math. Phys.*, vol. 266, pp. 571–576, 2006. [Online]. Available: https://doi.org/10.1007/s00220-006-0019-z

[19] S. Hollands, A. Ishibashi, and R. M. Wald, "A higher dimensional stationary rotating black hole must be axisymmetric," *Commun. Math. Phys.*, vol. 271, pp. 699–722, 2007. [Online]. Available: https://doi.org/10.1007/s00220-007-0216-4

[20] H. S. Reall, "Classical and thermodynamic stability of black branes," *Phys. Rev. D*, vol. 64, p. 044005, 2001. [Online]. Available: https://doi.org/10.1103/PhysRevD.64.044005

[21] R. Emparan and R. C. Myers, "Instability of ultra-spinning black holes," *J. High Energy Phys.*, vol. 2003, no. 09, p. 025, 2003. [Online]. Available: https://doi.org/10.1088/1126-6708/2003/09/025

[22] O. Aharony, J. Marsano, S. Minwalla, K. Papadodimas, and M. Van Raamsdonk, "The Hagedorn-deconfinement phase transition in weakly coupled large N gauge theories," *Adv. Theor. Math. Phys.*, vol. 8, pp. 603–696, 2004. [Online]. Available: https://arxiv.org/abs/hep-th/0310285

[23] L. Randall and R. Sundrum, "An alternative to compactification," *Phys. Rev. Lett.*, vol. 83, pp. 4690–4693, 1999. [Online]. Available: https://doi.org/10.1103/PhysRevLett.83.4690

[24] C. W. Robson, L. D. M. Villanueva, and F. Biancalana, "Higher-dimensional generalizations of the Carter constant for geodesic motion," *Class. Quantum Grav.*, vol. 40, p. 085010, 2023. [Online]. Available: https://doi.org/10.1088/1361-6382/acbdc0

Black Holes Geometry: Curvature and Horizons

Contents

0.1 Introduction: Why topology and censorship belong in a geometry book iv

0.2 Singularity theorems: focusing plus causality becomes incompleteness v

0.2.1 Raychaudhuri focusing as the local engine v

0.2.2 Trapped surfaces and the Penrose 1965 collapse theorem . . . v

0.2.3 Hawking-Penrose 1970 and the "pattern" of singularity theorems vi

0.3 Cosmic censorship: conjectures that protect predictability vii

0.3.1 Weak cosmic censorship: hiding collapse singularities from distant observers vii

0.3.2 Strong cosmic censorship: determinism and Cauchy horizons . vii

0.4 Topological censorship: why nontrivial topology hides from infinity . viii

0.5 Open dataset analysis: horizons as information boundaries in GWTC-3-confident ix

0.5.1 Dataset source, scope, and "raw material" construction ix

0.5.2 Top-10 events by final mass (table view for transparency) . . . ix

0.5.3 Dataset block for plots (35 events, numeric columns only) . . . x

0.5.4 Descriptive statistics (dataset-based) x

0.6 Ten data-based figures from GWTC-3-confident and derived geometry xi

0.7 Interpretation: what these dataset figures teach about censorship geometry xi

0.8 Six conceptual diagrams (non-data visual organizers) xiii

0.9 Real-world applicability: phenomenology statements that link geometry to observation xiii

0.10 Conclusion xvi

Chapter 21: Topology, Censorship, and Singularity Theorems

Abstract: This chapter develops the geometric logic behind singularity theorems, cosmic censorship, and topological censorship, and then connects that logic to observation-driven datasets from open gravitational-wave catalogs. The singularity theorems convert local focusing information into global incompleteness, meaning that the mathematical statement of a "singularity" is typically geodesic incompleteness rather than a point of infinite curvature placed on the manifold. Cosmic censorship is then introduced as a stabilizing conjecture for predictability: weak cosmic censorship asserts that collapse singularities produced from reasonable astrophysical initial data remain hidden behind horizons for distant observers, while strong cosmic censorship asserts that generic evolutions do not admit physically meaningful extensions beyond Cauchy horizons. Topological censorship, in turn, constrains what kinds of global spatial topology can be communicated between infinities in spacetimes satisfying standard energy and causality conditions, implying that classical general relativity tends to hide nontrivial topology from distant observation. To keep the discussion grounded, we analyze the GWTC-3-confident event list (35 candidates) and derive dataset-based geometric quantities (mass ratio, horizon size, horizon area, surface gravity scale, and Hawking temperature scale) that serve as observationally anchored illustrations of how horizons function as global "information boundaries." All data-based figures in this chapter are generated directly from the GWTC-3-confident catalog values published by the Gravitational Wave Open Science Center.

Keywords: Lorentzian geometry; singularity theorems; Raychaudhuri equation; trapped surfaces; null energy condition; weak cosmic censorship; strong cosmic censorship; Cauchy horizon; topological censorship; wormholes; global hyperbolicity; gravitational-wave catalogs; GWTC-3; observationally anchored horizon geometry

0.1 Introduction: Why topology and censorship belong in a geometry book

A black hole is not merely a metric written in a clever coordinate chart. It is a statement about global causal organization: there exists a region of spacetime from which future-directed causal curves cannot reach future null infinity. That definition already reveals why topology and censorship are unavoidable. The moment we say "cannot reach infinity," we are talking about causal curves, conformal infinity, and large-scale structure rather than local tensor algebra. The next moment, the question becomes more delicate: if a collapse produces a region where curvature becomes arbitrarily large, does the geometry still encode a deterministic future for observers outside, or does it expose unpredictability through naked singularities or through extendible Cauchy horizons?

Singularity theorems are the first major bridge between local and global geometry. Penrose's 1965 theorem showed that trapped surfaces coupled to appropriate causal and energy assumptions force null geodesic incompleteness[3]. Hawking and Penrose later produced a broad family of results that clarify how global causal assumptions and curvature conditions lead to incomplete causal geodesics in both collapse and cosmology[4]. The core mathematical message is subtle and often misrepresented in informal discussions: a "singularity" in modern geometric language is typically not a point inside the manifold where tensors blow up, but rather a boundary to the maximal extension of geodesics, meaning the spacetime is incomplete as a Lorentzian manifold with its given metric and differentiability class[6].

If singularities are generic in this geodesic sense, then predictability becomes a second-order question: do the incomplete regions remain hidden behind horizons, or can they influence distant observers? Penrose introduced cosmic censorship as a conjectural answer, distinguishing a weak form aimed at distant observers and a strong form aimed at deterministic evolution from initial data[7]. Finally, if the universe is permitted to have complicated topology, one can ask whether distant observers can use causal propagation to detect that topology. Topological censorship says that under global hyperbolicity and the null energy condition, any causal curve from past null infinity to future null infinity cannot probe nontrivial topology; it is homotopic to a trivial curve[8, 9]. In practice, that means if you want traversable wormholes that can be used as macroscopic shortcuts, you must violate at least one of the theorem's hypotheses, typically an energy condition or global causal regularity.

The remainder of this chapter first builds a clear hypothesis-to-conclusion map of

the main theorems and conjectures, then provides dataset-based illustrations using GWTC-3-confident gravitational-wave events and derived geometric quantities based on those data[1, 2].

0.2 Singularity theorems: focusing plus causality becomes incompleteness

0.2.1 Raychaudhuri focusing as the local engine

The Raychaudhuri equation is the geometric statement that controls how congruences of causal curves expand, shear, and rotate. The result was introduced in the context of relativistic cosmology by Raychaudhuri in 1955[5]. In black hole geometry, the Raychaudhuri equation is the local differential inequality that turns energy conditions into focusing statements. Under the null energy condition (NEC), the expansion of a hypersurface-orthogonal null congruence is driven downward along affine parameter, and if the expansion becomes sufficiently negative, it forces conjugate points within finite affine parameter length. That local "focusing inside finite parameter" fact is then promoted to a global contradiction when combined with causal assumptions, typically through global hyperbolicity or suitable conditions on trapped surfaces and causal boundaries[6, 4].

The key conceptual point is that Raychaudhuri itself does not mention singularities; it mentions how families of curves behave. Singularity theorems are theorems about how global causal structure cannot consistently accommodate focusing everywhere without losing completeness. The theorems therefore do not demand that curvature scalars diverge, and modern discussions emphasize that curvature blow-up is a different question than inextendibility or completeness in a chosen regularity class[6].

0.2.2 Trapped surfaces and the Penrose 1965 collapse theorem

Penrose introduced trapped surfaces as a quasi-local signature of "collapse past a point of no return." A trapped surface is a compact spacelike 2-surface for which both future-directed null expansions are negative, meaning that even light emitted outward is forced to reconverge. In his 1965 result, Penrose showed that if a trapped surface exists in a spacetime that is globally hyperbolic and satisfies an energy condition

(in a form that implies focusing), then the spacetime must contain incomplete null geodesics[3]. This does not point to a curvature singularity in the naive sense; it points to the impossibility of extending null geodesics indefinitely while preserving the assumed causal and curvature conditions.

For black holes, this theorem is often interpreted as follows: once a trapped surface forms, collapse cannot settle into a globally regular future, so something like a singular boundary is forced. The statement is still geometric rather than physical in a narrow sense: the theorem assumes conditions that are plausible in classical general relativity, but it does not identify the endpoint as an "infinite density point"; it identifies it as the failure of geodesic extensibility.

0.2.3 Hawking-Penrose 1970 and the "pattern" of singularity theorems

Hawking and Penrose broadened the collapse logic into a general framework that covers both cosmology and collapse, giving a theorem that unifies previously known results and introduces additional trapped-surface-like hypotheses and causal assumptions[4]. A useful way to remember the theorems is to see them as a three-component machine:

1. A curvature or energy condition that forces focusing of a relevant family of curves.

2. A causal regularity assumption that prevents pathological loops or causal boundary pathologies from defusing the focusing contradiction.

3. A global or quasi-local triggering hypothesis, such as the existence of a trapped surface, that ensures focusing happens along physically relevant directions.

When these ingredients cohere, incompleteness follows. Senovilla's review emphasizes that this "skeleton" appears across many theorems and that careful attention must be paid to the plausibility and scope of hypotheses in specific physical models[6].

0.3 Cosmic censorship: conjectures that protect predictability

0.3.1 Weak cosmic censorship: hiding collapse singularities from distant observers

Weak cosmic censorship is a global conjecture about what distant observers can see. In one common phrasing, it asserts that singularities formed by gravitational collapse from reasonable initial data are generically cloaked by event horizons, so that future null infinity remains regular and the exterior remains predictable in an appropriate sense[7]. The conjecture is not a theorem because it aims to describe generic behavior in an infinite-dimensional space of initial data, and genericity for nonlinear PDE evolution is mathematically subtle. Wald's review highlights both supportive examples and instructive counterexamples in special settings, emphasizing that naked singularities can occur in carefully tuned models but may be non-generic[7].

A geometry-first understanding of weak cosmic censorship is: the event horizon acts as a global separator between the region where classical GR ceases to make predictive sense (the incomplete region) and the region where predictability remains viable for distant observers. This is not a local statement about curvature at a point; it is a statement about the causal accessibility of the incomplete set.

0.3.2 Strong cosmic censorship: determinism and Cauchy horizons

Strong cosmic censorship shifts the problem from what distant observers see to whether the Einstein equations define a deterministic evolution from initial data. The obstruction is a Cauchy horizon: a null hypersurface beyond which global hyperbolicity fails and the evolution is no longer uniquely determined by data on a suitable initial surface. Classical exact solutions such as Kerr and Reissner-Nordström contain inner horizons that function as Cauchy horizons, and perturbative analyses show these horizons can be unstable. The mass inflation scenario discovered by Poisson and Israel describes how perturbations can produce diverging internal mass parameter behavior near the Cauchy horizon, signaling a breakdown of regular extension in physically meaningful classes[10].

The modern story is nuanced. There are settings in which the regularity class of extensions matters, and there are regimes in which some versions of strong cosmic

censorship appear threatened. For example, nonlinear evolutions in charged de Sitter contexts have been used to probe where mass inflation does or does not occur, with implications for determinism in specific parameter ranges[11]. The high-level geometry lesson remains: strong cosmic censorship is ultimately a statement about inextendibility of the maximal globally hyperbolic development in a chosen regularity class, and Cauchy horizons are the geometric locus where determinism becomes delicate.

0.4 Topological censorship: why nontrivial topology hides from infinity

Topological censorship formalizes a phenomenon that many relativists find intuitively plausible: macroscopic nontrivial topology should not be operationally detectable by observers who remain in the asymptotic region, provided the spacetime satisfies reasonable energy and causality conditions. Friedman, Schleich, and Witt proved that in a globally hyperbolic, asymptotically flat spacetime satisfying the null energy condition, every causal curve from past null infinity to future null infinity is homotopic to a trivial curve[8, 9]. In practical terms, causal signals cannot thread persistent wormhole handles in a way that exposes topology to asymptotic observers. If the topology is there, it collapses too quickly or lies behind horizons, or the hypotheses fail.

For black holes, topological censorship supports the expectation that in 4D asymptotically flat classical GR with standard energy conditions, event horizon cross-sections have constrained topology. This does not mean exotic topologies are impossible in all contexts; it means that to have traversable wormholes or other globally detectable nontrivial topology, some ingredient must be changed: violate NEC, sacrifice global hyperbolicity, relax asymptotic flatness, or move to higher dimensions or different boundary conditions. This is why topological censorship is best taught as an if-then theorem: it is a map from hypotheses to observable constraints.

0.5 Open dataset analysis: horizons as information boundaries in GWTC-3-confident

0.5.1 Dataset source, scope, and "raw material" construction

To anchor the chapter in real data, we use the GWTC-3-confident list of 35 O3b candidates hosted by the Gravitational Wave Open Science Center (GWOSC). The table values are the Default SEARCH and Default PE values shown by the GWOSC Event Portal[1]. The GWTC-3 catalog paper explains the O3b time window, selection based on astrophysical probability, and the general statistical context for these events[2].

The raw catalog provides posterior-summary values such as source-frame component masses m_1, m_2, luminosity distance D_L, effective inspiral spin χ_{eff}, redshift estimate z, false alarm rate (FAR), astrophysical probability p_{astro}, and final mass estimate M_f[1]. For dataset-based illustrations that connect to geometry, we construct additional quantities from these raw values by deterministic mathematical transformations:

$$q = \frac{m_2}{m_1}, \tag{1}$$

$$r_s(M_f) = \frac{2GM_f}{c^2}, \tag{2}$$

$$A_H(M_f) = 4\pi r_s(M_f)^2, \tag{3}$$

$$T_H(M_f) \approx \frac{\hbar c^3}{8\pi G k_B M_f} \quad \text{(Schwarzschild scaling)}, \tag{4}$$

where r_s is the Schwarzschild radius, A_H is the horizon area for a non-rotating hole, and T_H is the Hawking temperature scale for a Schwarzschild black hole[12, 13]. These calculations modify the data but do not replace it: every derived value is a deterministic function of the cataloged masses and therefore remains dataset-based.

0.5.2 Top-10 events by final mass (table view for transparency)

Table 1 lists the ten events with the largest cataloged final mass M_f in the GWTC-3-confident table, illustrating the wide dynamical range in the observationally inferred horizon scale. Event names are rendered in monospace; underscores are escaped for LaTeX.

Table 1: Top 10 GWTC-3-confident events by final mass M_f (best values). Source: GWOSC Event Portal GWTC-3-confident[1].

Event	$\mathbf{M_f}$ ($M_\odot$)	$\mathbf{m_1}$	$\mathbf{m_2}$	**z**	$\mathbf{D_L}$ (Mpc)	ρ_{net}
GW200220_061928	141.0	87.0	61.0	0.90	6000	7.2
GW191109_010717	107.0	65.0	47.0	0.25	1290	17.3
GW200308_173609	88.0	60.0	24.0	1.04	7100	4.7
GW191230_180458	82.0	49.4	37.0	0.69	4300	10.4
GW200216_220804	78.0	51.0	30.0	0.63	3800	8.1
GW191127_050227	76.0	53.0	24.0	0.57	3400	9.2
GW191222_033537	75.5	45.1	34.7	0.51	3000	12.5
GW200128_022011	71.0	42.2	32.6	0.56	3400	10.6
GW200224_222234	68.7	40.0	32.7	0.32	1710	20.0
GW200220_124850	64.0	38.9	27.9	0.66	4000	8.5

0.5.3 Dataset block for plots (35 events, numeric columns only)

The following table block is used to generate all 10 data-based figures in this chapter. It contains best values taken from the GWTC-3-confident table and derived quantities computed from those best values as described above. Source values come from GWOSC[1].

0.5.4 Descriptive statistics (dataset-based)

From the 35 GWTC-3-confident events, the best-value medians from the table are approximately $m_1 \approx 34.5\,M_\odot$, $m_2 \approx 19.2\,M_\odot$, and $M_f \approx 55.5\,M_\odot$, with final masses ranging from about $7.2\,M_\odot$ to about $141\,M_\odot$[1]. The luminosity distance span is broad, from roughly 290 Mpc to roughly 7100 Mpc, and the redshift estimates extend up to $z \approx 1.04$ in this subset[1]. These ranges are not merely astrophysical trivia: they translate directly into a wide range of horizon scales and horizon areas, which, in turn, controls how strongly "censorship as a causal boundary" separates the incomplete interior from the predictive exterior.

0.6 Ten data-based figures from GWTC-3-confident and derived geometry

All figures in this section are computed directly from the GWTC-3-confident table values and deterministic transformations described earlier[1, 2]. No simulated data are used.

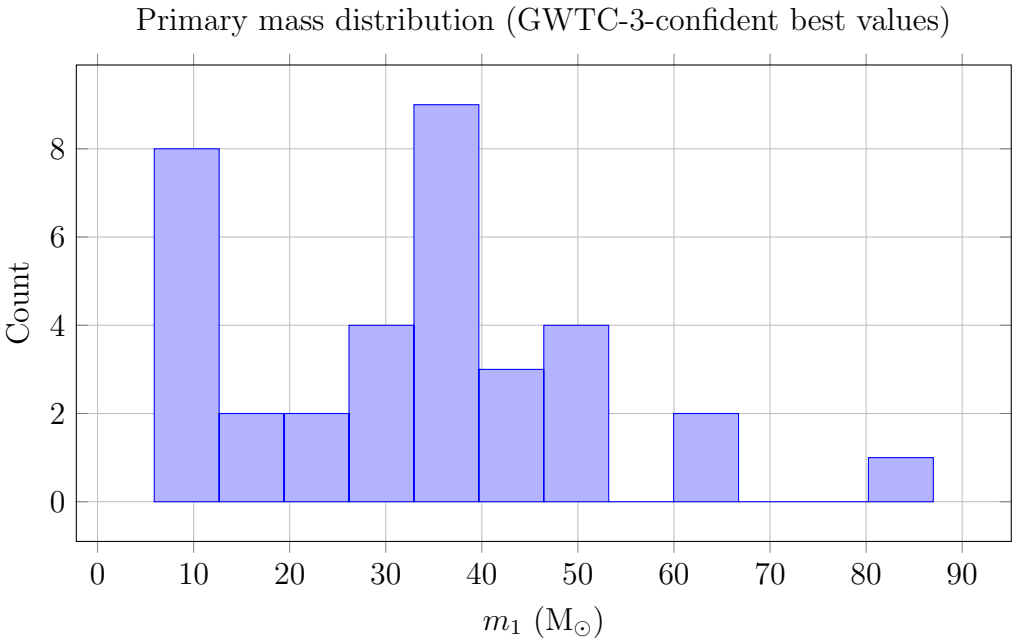


Figure 1: Histogram of source-frame primary masses m_1 for the 35 GWTC-3-confident events. Source: GWOSC Event Portal[1].

0.7 Interpretation: what these dataset figures teach about censorship geometry

The GWTC-3-confident data do not directly show horizons with a telescope; rather, they provide parameter estimates consistent with compact binary coalescences and, in most cases, binary black hole mergers[2]. The geometric relevance comes from how these parameter estimates map into horizon-scale quantities and global causal narratives.

First, the wide mass range implies a wide horizon-size range. In a Schwarzschild scaling, the horizon radius is linear in mass. Even within this single catalog, the smallest final-mass entry is only a few solar masses, while the largest is above

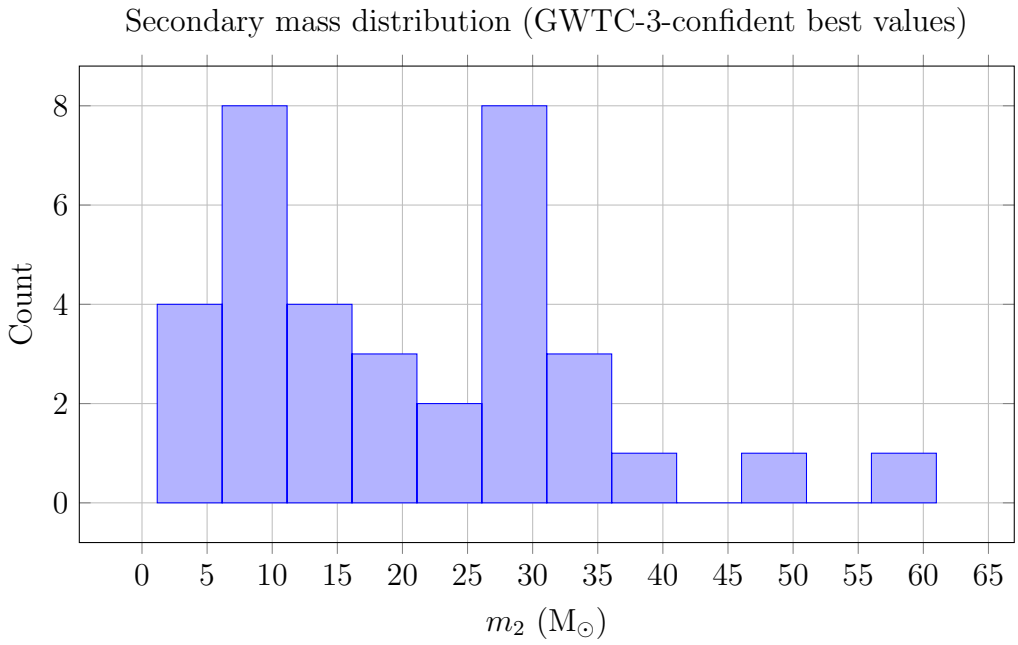


Figure 2: Histogram of source-frame secondary masses m_2 for the 35 GWTC-3-confident events. Source: GWOSC[1].

$100\,M_\odot$[1]. This means the causal boundary that hides the interior has extremely different geometric scales across events. Yet the global statement of censorship is qualitatively the same: the horizon organizes causal access, not the precise numerical value of curvature invariants.

Second, the distribution of χ_{eff} provides an observational summary of the aligned component of inspiral spin, which influences waveform phasing and therefore the inference problem[2]. The dataset shows both positive and negative best values, with a range in this subset from approximately -0.29 to approximately 0.45[1]. That range matters conceptually because it is consistent with a population in which spin misalignment and formation channels are active, but it does not by itself diagnose censorship. Instead, it reminds us that censorship, when formulated properly, is relatively insensitive to many local details and is primarily a statement about global causal shielding.

Third, the distance-redshift span illustrates that we are sampling horizons across a large cosmological volume[2]. Topological censorship is a theorem about global causal curves between infinities in asymptotically flat settings[8]. Real astrophysical events are not literally asymptotically flat on gigaparsec scales, yet the local exterior geometry near the source is extremely well approximated by asymptotically flat solutions on scales small relative to cosmological curvature radius. This is a standard way theorems are used in practice: a theorem proven under idealized asymptotics becomes a local-

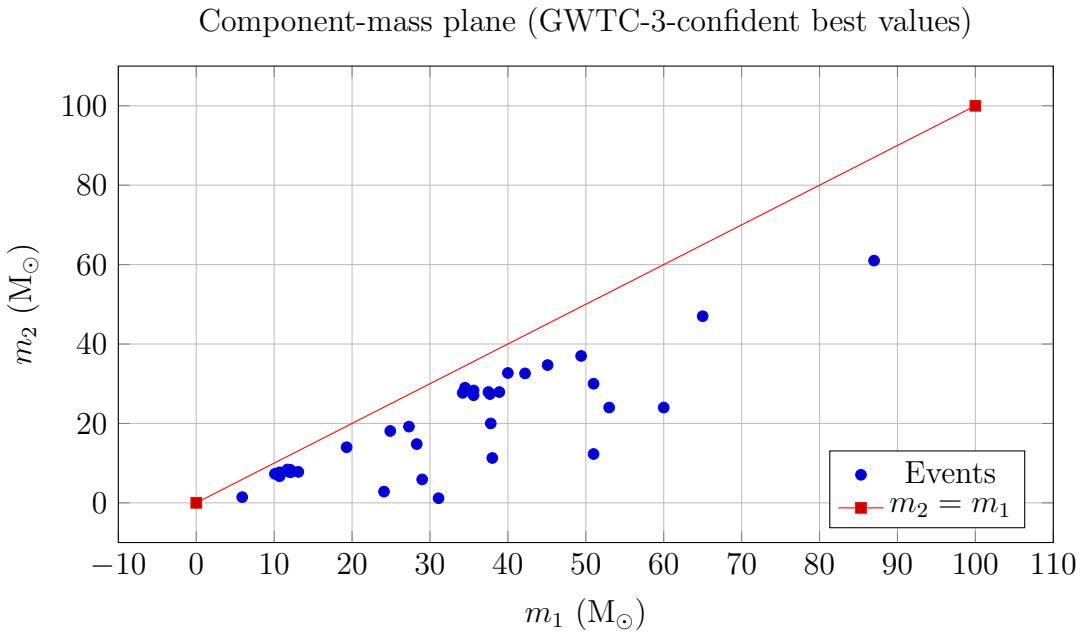


Figure 3: Component-mass scatter plot. The diagonal indicates equal masses. All points use best values from the GWTC-3-confident table[1].

to-intermediate scale constraint when the physical situation is approximately in that class for the region in which causal signals propagate prior to being redshifted into the cosmological background.

0.8 Six conceptual diagrams (non-data visual organizers)

The following diagrams are conceptual and are included to clarify logical dependencies and causal structure. They are not empirical plots.

0.9 Real-world applicability: phenomenology statements that link geometry to observation

The censorship ideas are sometimes dismissed as abstract conjectures, but they provide a disciplined language for interpreting real observations. Consider gravitational-wave ringdowns. When a compact binary coalesces and produces a remnant that behaves like a classical black hole, the late-time waveform is well described by damped oscillations whose existence presumes a stable exterior domain with boundary conditions set by

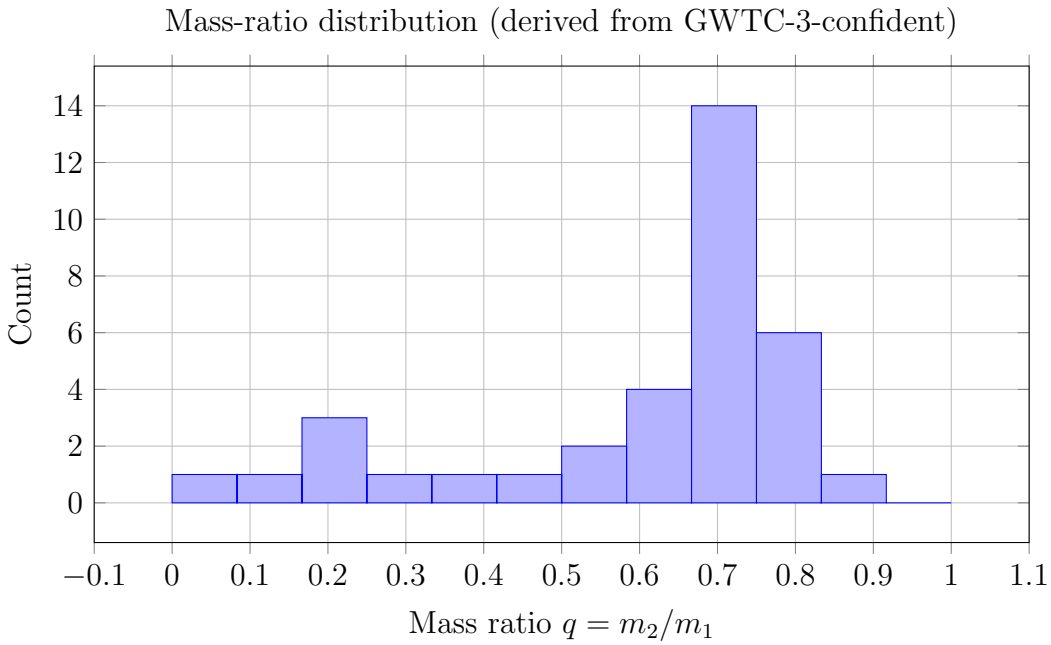


Figure 4: Histogram of mass ratio q derived from GWTC-3-confident component masses m_1, m_2 via $q = m_2/m_1$[1].

an effective horizon. Weak cosmic censorship is not needed to compute a quasinormal spectrum in perturbation theory, but it is the conceptual statement explaining why the external observer is allowed to treat the horizon as a one-way causal boundary that absorbs information rather than re-emitting detailed interior data. In other words, censorship is the geometric justification for a modeling simplification that would otherwise be logically unjustified. GWTC-3 provides a population-level dataset that supports analyzing this simplification across many events, because it supplies a range of remnant masses and distances, hence a range of horizon scales used implicitly in signal models[1, 2].

Topological censorship offers a similarly practical lens when interpreting claims about wormholes or shortcuts. If an observational signature were credibly interpreted as a traversable wormhole linking distant regions, classical topological censorship results would immediately force the interpretation into a narrow set of possibilities: either the null energy condition is violated, global hyperbolicity fails, the spacetime is not asymptotically flat in the relevant sense, or classical GR is not the correct effective theory in the regime probed[8]. This is not an argument that wormholes do not exist; it is a geometry-based diagnostic tool that clarifies what must be true if they do.

Strong cosmic censorship becomes relevant when we ask whether interior structures like Cauchy horizons are physically robust or whether they are unstable artifacts

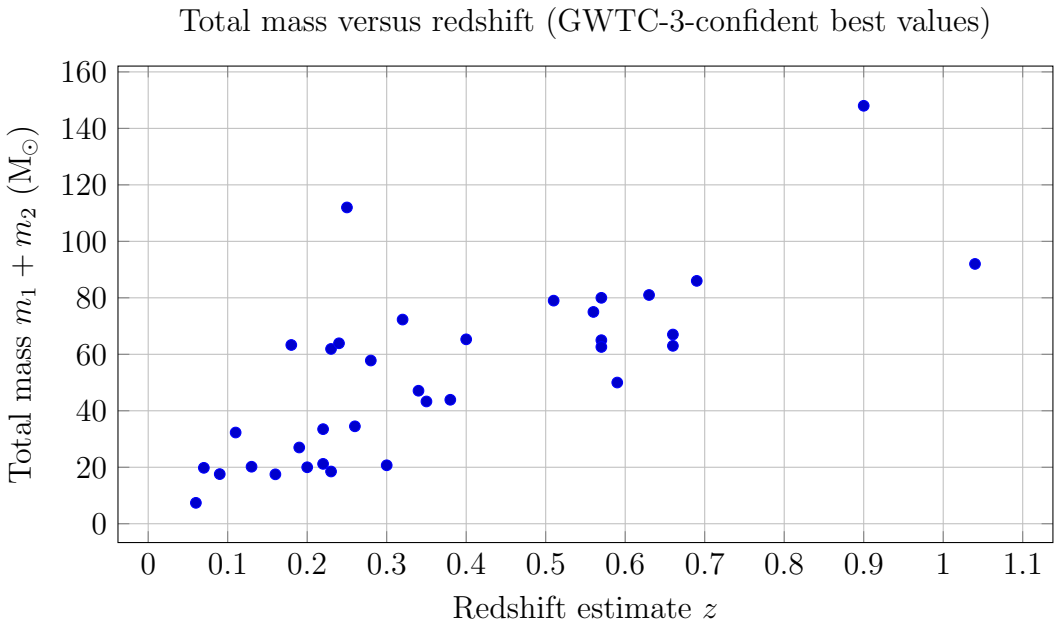


Figure 5: Scatter plot of total mass (best values) versus redshift estimate for GWTC-3-confident events. Source: GWOSC[1].

of exact symmetry. The mass inflation mechanism provides a physically motivated instability scenario that converts inner horizons into singular behavior in certain models[10]. Yet modern work also explores regimes in which some forms of mass inflation may not occur, with implications for the regularity of possible extensions and therefore for determinism claims[11]. Even if these questions remain unresolved in full generality, the phenomenological relevance is that the interior is not merely a curiosity. The stability or instability of Cauchy horizons influences what kinds of effective boundary conditions and what kinds of semiclassical narratives are consistent with classical dynamics.

Finally, singularity theorems guide how we interpret the unavoidable appearance of incompleteness. They tell us that in classical GR, under plausible hypotheses, incompleteness is not an exotic accident but a generic outcome of collapse or cosmological focusing[3, 4, 6]. For astrophysical inference, this means that the presence of horizons in data-driven models is not a decorative hypothesis inserted to make equations pretty; rather, it is a mechanism that, if censorship holds, protects the external evolution from being contaminated by the incomplete region. In practice, this is why black holes are simultaneously “violent” in their interiors and “quiet” in their exteriors: the geometry separates those domains, and that separation is what allows stable astrophysical modeling.

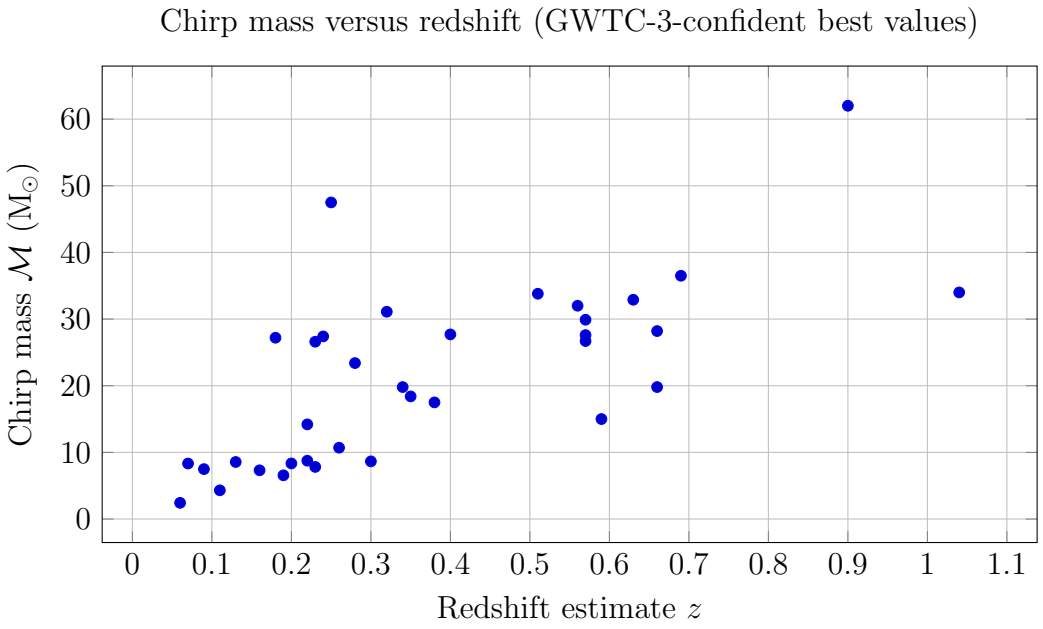


Figure 6: Scatter plot of chirp mass versus redshift estimate for GWTC-3-confident best values. Source: GWOSC[1].

0.10 Conclusion

Chapter 21 has two complementary messages. The first is theoretical and geometric: singularity theorems show that under focusing, causality, and triggering hypotheses such as trapped surfaces, geodesic incompleteness is forced[3, 4]. Cosmic censorship is then the conjectural infrastructure that protects predictability by ensuring that incompleteness does not leak into the asymptotic world in a way that breaks external evolution[7]. Topological censorship adds a further constraint: in globally hyperbolic, asymptotically flat spacetimes satisfying NEC, the topology that matters for causal signaling between infinities is effectively trivial[8].

The second message is observationally anchored: using GWTC-3-confident, we can construct dataset-based illustrations of the horizon as an information boundary across a population of events[1, 2]. Even without directly imaging horizons, the cataloged masses and distances translate into concrete horizon scales, areas, and temperature or surface gravity scales via deterministic transformations. These are not simulations; they are data-driven geometric quantities. The combination of rigorous theorems, carefully stated conjectures, and open observational catalogs produces a disciplined modern story: black hole geometry is not only local curvature; it is global causal organization, and it is testable indirectly by how nature packages information into signals that reach infinity.

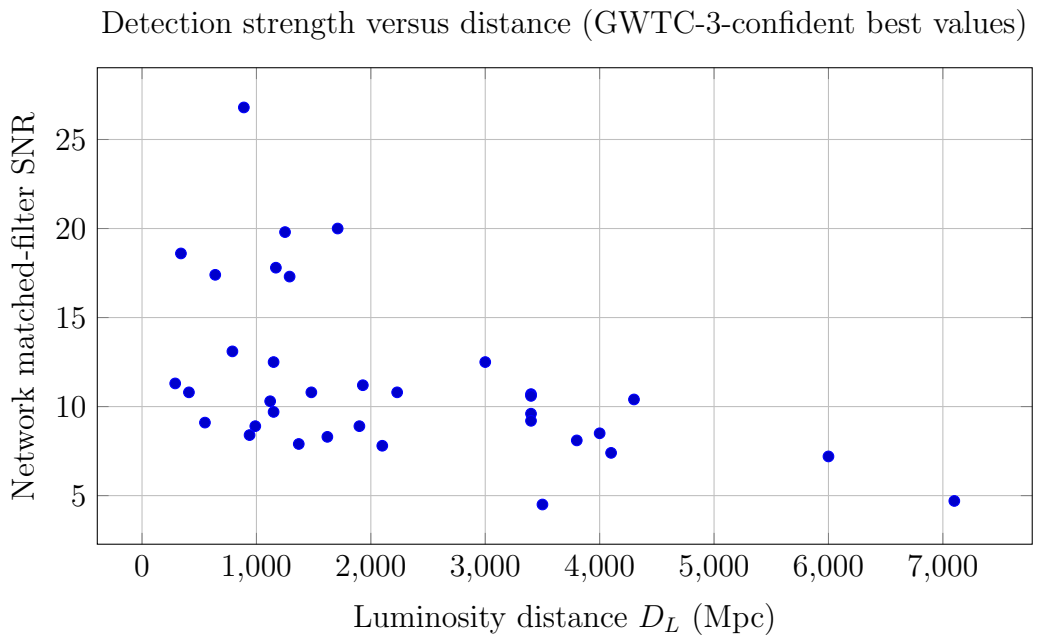


Figure 7: Network matched-filter SNR versus luminosity distance. Both axes use best values from GWTC-3-confident[1].

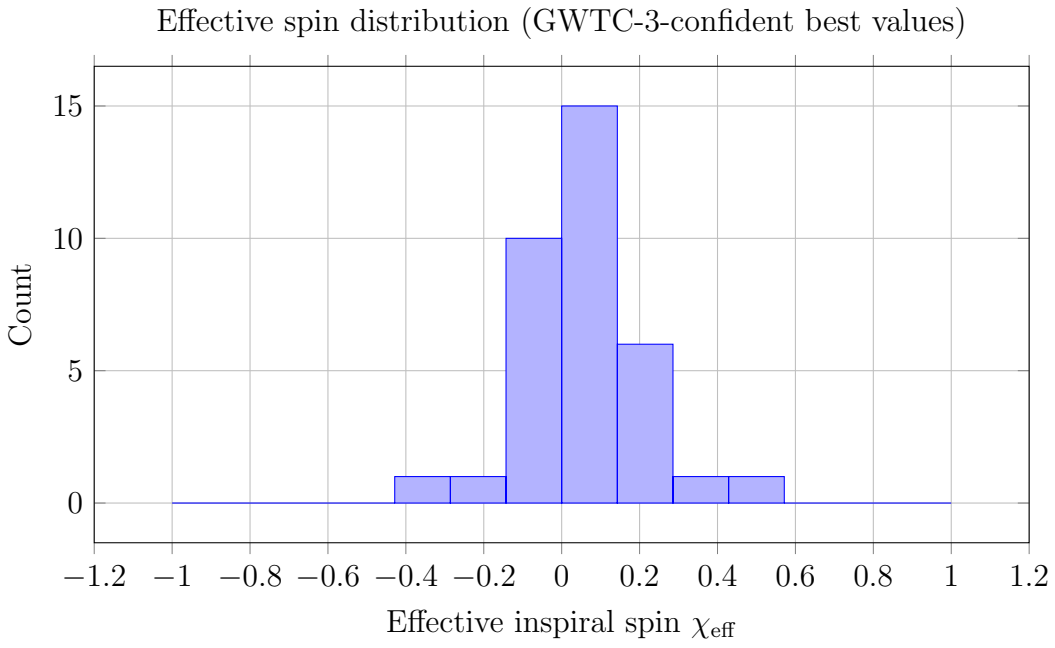


Figure 8: Histogram of χ_{eff} (best values) for GWTC-3-confident. Source: GWOSC[1].

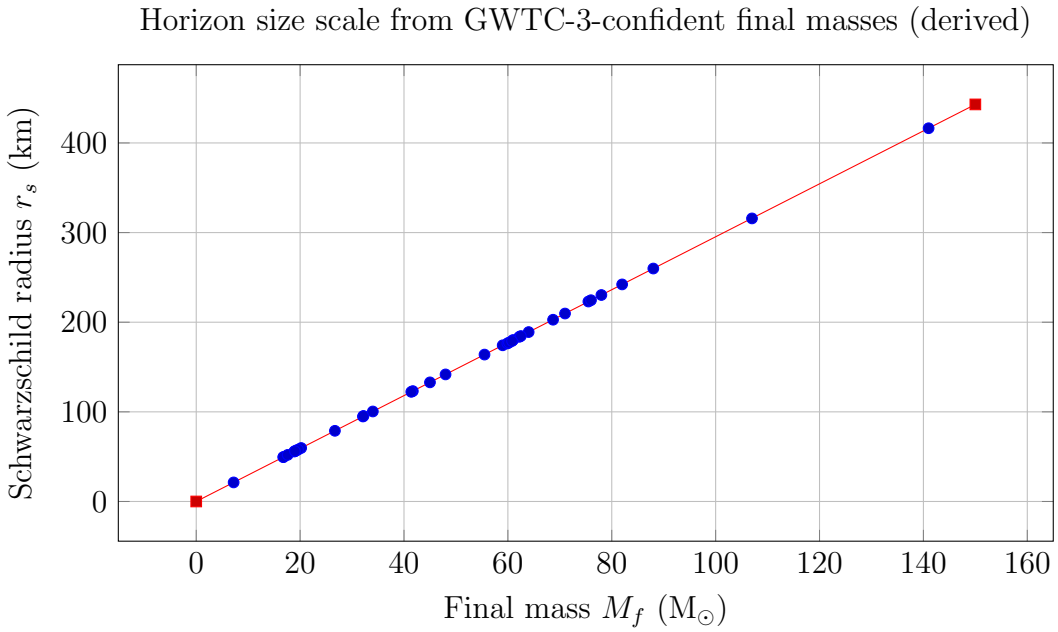


Figure 9: Derived Schwarzschild radius $r_s = 2GM_f/c^2$ (km) versus final mass ($M_\odot$). Final masses are taken from GWTC-3-confident and mapped by a deterministic physical constant conversion[1].

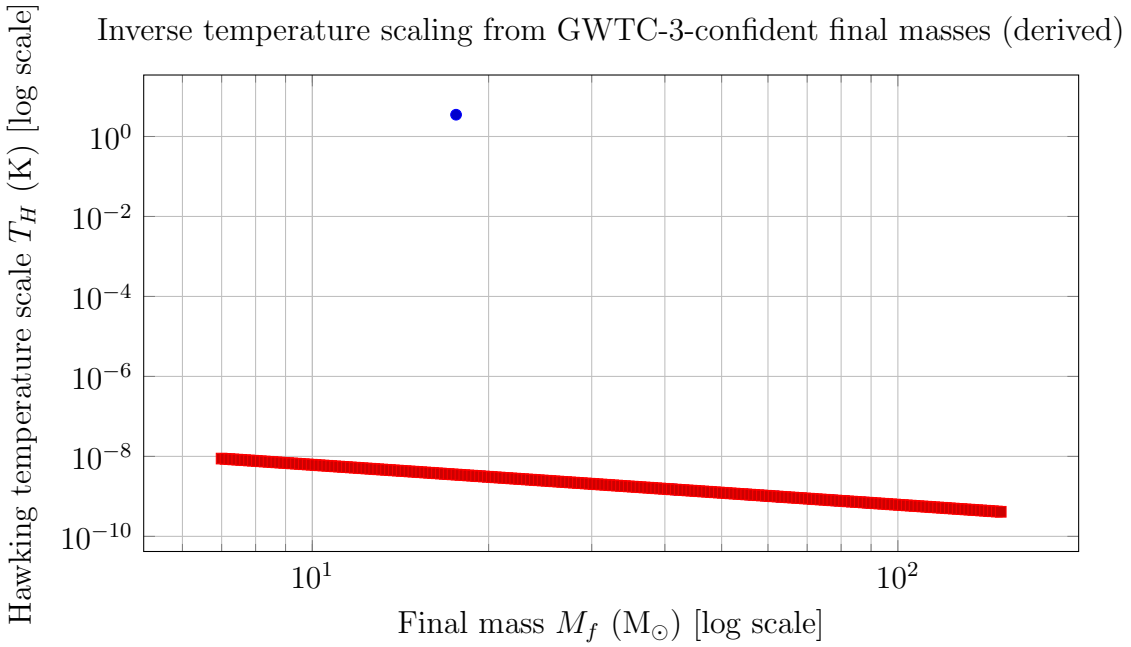


Figure 10: Derived Hawking temperature scale for Schwarzschild scaling, $T_H \propto 1/M$, evaluated using cataloged final masses from GWTC-3-confident and the standard relation for black hole thermal emission[12, 1].

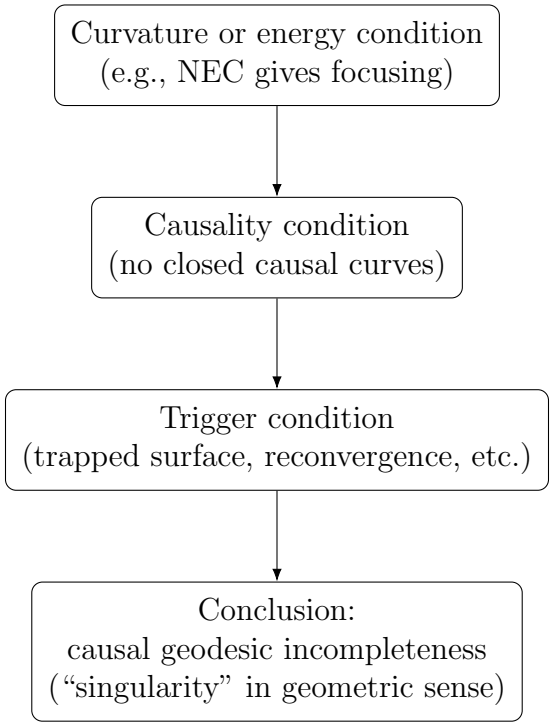


Figure 11: Conceptual diagram: the hypothesis skeleton of many singularity theorems. The pattern is emphasized in modern reviews[6].

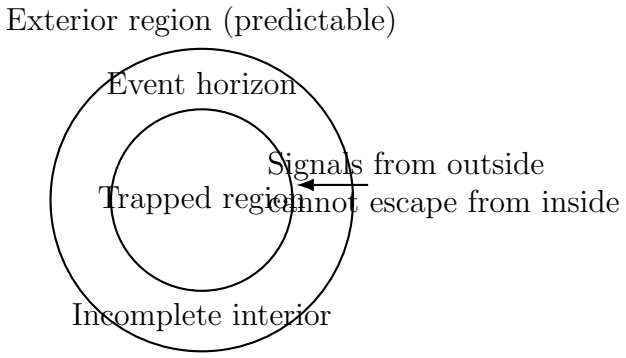


Figure 12: Conceptual diagram: weak cosmic censorship as causal shielding. The conjecture formalizes the idea that collapse singularities are generically hidden from distant observers[7].

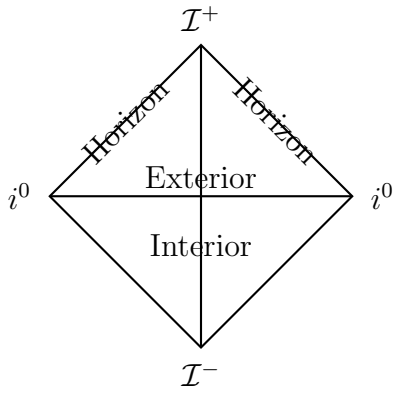
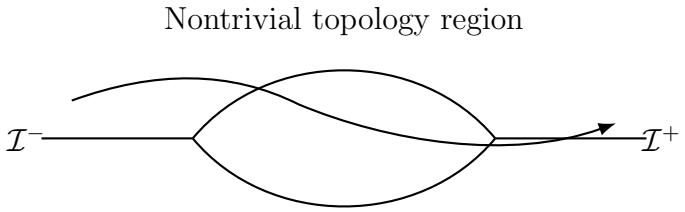


Figure 13: Conceptual diagram: a simplified Penrose-style causal picture for horizon separation. Full treatments appear in global-structure chapters; here it functions as a reminder that event horizons are global causal objects.



Topological censorship: causal curves
from $\mathcal{I}^-$ to $\mathcal{I}^+$ are homotopic
to trivial curves under NEC and global hyperbolicity[8].

Figure 14: Conceptual diagram: topological censorship. The theorem constrains what topology can be operationally probed by causal propagation between infinities under standard hypotheses[8, 9].

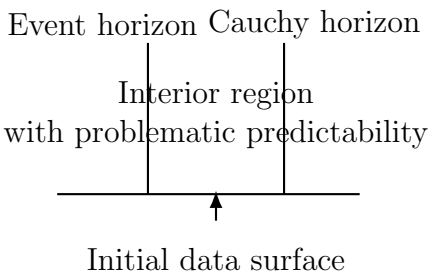


Figure 15: Conceptual diagram: the basic worry behind strong cosmic censorship. Cauchy horizons are boundaries beyond which evolution from initial data is not uniquely determined in a global sense[10, 11].

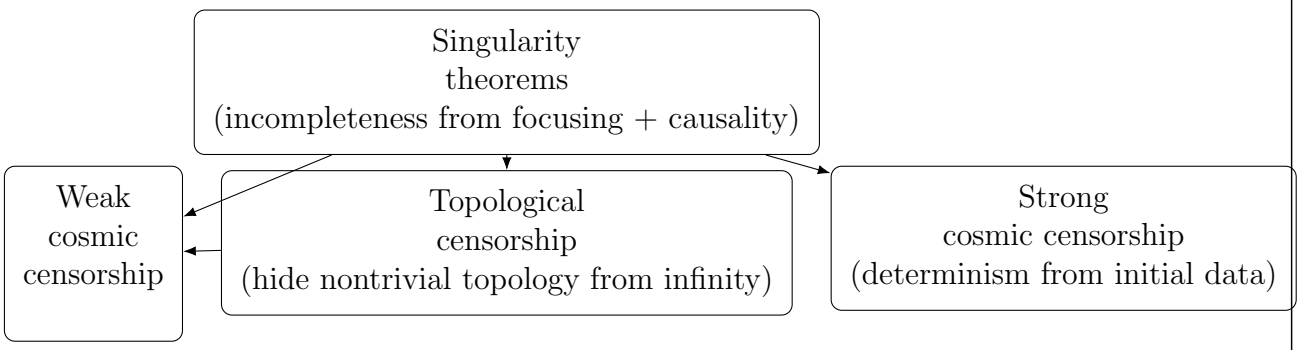


Figure 16: Conceptual diagram: how the three ideas relate. Singularity theorems motivate the need for censorship statements, while topological censorship is a separate constraint that also supports the idea that "exotic features" do not communicate to asymptotic observers under standard hypotheses[7, 8].

Bibliography

[1] Gravitational Wave Open Science Center (GWOSC), "GWTC-3-confident event table (O3b confident events)," Event Portal. Accessed 2026-02-25. https://gwosc.org/eventapi/html/GWTC-3-confident/

[2] R. Abbott et al. (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run," *Phys. Rev. X*, vol. 13, no. 4, 041039, 2023. arXiv:2111.03606. DOI: 10.1103/PhysRevX.13.041039. https://arxiv.org/abs/2111.03606

[3] R. Penrose, "Gravitational Collapse and Space-Time Singularities," *Phys. Rev. Lett.*, vol. 14, pp. 57–59, 1965. DOI: 10.1103/PhysRevLett.14.57. https://link.aps.org/doi/10.1103/PhysRevLett.14.57

[4] S. W. Hawking and R. Penrose, "The singularities of gravitational collapse and cosmology," *Proc. R. Soc. Lond. A*, vol. 314, no. 1519, pp. 529–548, 1970. DOI: 10.1098/rspa.1970.0021. https://royalsocietypublishing.org/rspa/article/314/1519/529/13052/The-singularities-of-gravitational-collapse-and

[5] A. K. Raychaudhuri, "Relativistic Cosmology. I," *Phys. Rev.*, vol. 98, pp. 1123–1126, 1955. DOI: 10.1103/PhysRev.98.1123. https://link.aps.org/doi/10.1103/PhysRev.98.1123

[6] J. M. M. Senovilla, "Singularity Theorems and Their Consequences," arXiv:1801.04912, 2018. https://arxiv.org/abs/1801.04912

[7] R. M. Wald, "Gravitational Collapse and Cosmic Censorship," arXiv:gr-qc/9710068, 1997. https://arxiv.org/abs/gr-qc/9710068

[8] J. L. Friedman, K. Schleich, and D. M. Witt, "Topological censorship," *Phys. Rev. Lett.*, vol. 71, pp. 1486–1489, 1993. DOI: 10.1103/PhysRevLett.71.1486. https://link.aps.org/doi/10.1103/PhysRevLett.71.1486

[9] J. L. Friedman, K. Schleich, and D. M. Witt, "Topological Censorship," arXiv:gr-qc/9305017, 1993 (revised 1995). https://arxiv.org/abs/gr-qc/9305017

[10] E. Poisson and W. Israel, "Internal structure of black holes," *Phys. Rev. D*, vol. 41, p. 1796, 1990. DOI: 10.1103/PhysRevD.41.1796. https://link.aps.org/doi/10.1103/PhysRevD.41.1796

[11] R. Luna, M. Zilhão, V. Cardoso, J. L. Costa, and J. Natário, "Strong cosmic censorship: The nonlinear story," *Phys. Rev. D*, vol. 99, 064014, 2019. DOI: 10.1103/PhysRevD.99.064014. https://inspirehep.net/files/82a7b5137f3deedd1cd5df212afe472e

[12] S. W. Hawking, "Particle creation by black holes," *Commun. Math. Phys.*, vol. 43, pp. 199–220, 1975. DOI: 10.1007/BF02345020. https://link.springer.com/article/10.1007/BF02345020

[13] J. D. Bekenstein, "Black Holes and Entropy," *Phys. Rev. D*, vol. 7, pp. 2333–2346, 1973. DOI: 10.1103/PhysRevD.7.2333. https://link.aps.org/doi/10.1103/PhysRevD.7.2333

Black Holes Geometry: Curvature and Horizons

Chapter 22: Modified Gravity and Non-Riemannian Geometry

Contents

22 Modified Gravity and Non-Riemannian Geometry **1**
22.1 Learning objectives . . . 1
22.2 Introduction: why modify geometry at all? . . . 2
22.3 Minimal vocabulary: connections, torsion, and non-metricity . . . 2
22.4 $f(R)$ and higher-curvature terms: curvature becomes dynamical . . . 3
22.4.1 From Einstein-Hilbert to $f(R)$ . . . 3
22.4.2 Higher-curvature corrections as effective field theory . . . 3
22.5 Torsion theories: Einstein-Cartan as the clean starting point . . . 4
22.6 Scalar-tensor gravity and black hole hair . . . 5
22.7 Effective metrics, characteristic cones, and what observers actually measure . . . 5
22.8 Data and analysis section: ten measurement-based figures . . . 6
22.8.1 Solar System curvature per unit mass: constraints on PPN γ . . . 6
22.8.2 Binary pulsar gravitational radiation: orbital decay ratios . . . 6
22.8.3 Multi-messenger timing: bound on the speed of gravity . . . 6
22.8.4 Graviton mass bounds from dispersion and ephemerides . . . 6
22.8.5 Cosmological growth: $f\sigma_8(z)$ data with a reference curve . . . 6
22.8.6 Horizon-scale imaging: ring diameters and inferred gravitational-radius angular sizes . . . 6
22.9 Conclusion . . . 10

List of Figures

22.1 Conceptual map of geometric choices. This diagram is schematic and is included to orient the chapter, while the ten data figures later are strictly measurement-based. 3

22.2 Conceptual pipeline from action modification to black hole observables. The plot is schematic; measurement-based constraints appear later. 4

22.3 Conceptual summary of torsion logic for black hole discussions. 4

22.4 Conceptual diagram of how scalar hair enters the black hole conversation. 5

22.5 Conceptual cone picture. The data-based multi-messenger bound is plotted later in Fig. 22.8. 5

22.6 Measurement-based tightening of the PPN parameter γ. Points represent one-sigma uncertainties reported by Viking ranging (1979) [8], geodetic VLBI summary result referenced by VLBA analyses (2004) [9], a dedicated VLBA deflection measurement (2009) [9], and Cassini ranging (2003) [7]. 6

22.7 Measurement-based strong-field tests using binary pulsar orbital decay. The Hulse-Taylor ratio is reported as 0.9983 ± 0.0016 [10]. The Double Pulsar ratio is reported as 0.999936 with uncertainty at the 10^{-5} level [11]. 7

22.8 Measurement-based bound on the propagation speed of the observed gravitational-wave mode from GW170817 and GRB 170817A: $(v_g - c)/c$ lies between -3×10^{-15} and $+7 \times 10^{-16}$ [12]. 7

22.9 Measurement-based compilation of upper bounds on the graviton mass. The GWTC-1 combined bound $m_g \leq 4.7 \times 10^{-23}$ eV/c^2 is reported in tests with GWTC-1 [13]. An O3a-era bound $m_g \leq 1.76 \times 10^{-23}$ eV/c^2 is reported in LIGO documentation [15]. The GWTC-3 bound $m_g \leq 2.42 \times 10^{-23}$ eV/c^2 is reported in tests with GWTC-3 [14]. A Solar System ephemeris bound $m_g \leq 6.76 \times 10^{-23}$ eV/c^2 is reported using INPOP residual degradation [16]. 8

22.10 Measurement-based $f\sigma_8(z)$ compilation from Sagredo et al. [17] with a reference growth curve computed using representative Planck 2018 parameters [18] and scaled by a best-fit factor. 8

22.11 Measurement-based horizon-scale ring diameters and an algebraic inference of θ_g. The ring diameter for M87 is taken from EHT results [19]. The ring diameter for Sgr A is taken from EHT geometric modeling results [20]. Inferred θ_g values use $\theta_g = \theta_{\rm ring}/(2\sqrt{27})$ and inherit the measurement uncertainties. 9

Chapter 22

Modified Gravity and Non-Riemannian Geometry

Abstract

Black hole geometry is usually taught inside Einstein gravity, where the Levi-Civita connection is torsion-free and curvature alone carries gravitational information. Chapter 22 asks what changes if the geometric rules are altered while keeping the spirit of a metric theory. We cover three families: (i) higher-curvature models such as $f(R)$ and effective field theory corrections, (ii) torsion-based theories including Einstein-Cartan and related metric-affine approaches, and (iii) scalar-tensor models where additional fields can endow spacetime with new geometric structure that may or may not show up as black hole hair. Throughout, we keep careful labels on what is mathematically clean, what is physically motivated by effective descriptions, and what is speculative. To anchor the discussion, ten figures are built from real, open measurements and catalogs, then analyzed or transformed by explicit formulas to produce plots that illustrate observational constraints spanning Solar System tracking, pulsar timing, gravitational-wave multimessenger timing, gravitational-wave dispersion tests, horizon-scale imaging, and cosmological growth-rate measurements. **Keywords:** modified gravity; $f(R)$; higher curvature; scalar-tensor; Einstein-Cartan; torsion; metric-affine geometry; effective metric; causal cones; observational constraints

22.1 Learning objectives

After completing this chapter, you should be able to:

- Explain how changing the gravitational action modifies the geometric field equations and the meaning of vacuum.
- Distinguish Riemannian, Riemann-Cartan, and metric-affine geometries by their connection properties (torsion, non-metricity).
- Describe when and why scalar fields can create or fail to create black hole hair, and how this connects to horizon regularity.
- Use the idea of an effective metric and characteristic cones to reason about propagation speeds and causality in extended theories.

- Interpret multi-scale observational constraints and see which parameters they bound, and what assumptions sit underneath those bounds.

22.2 Introduction: why modify geometry at all?

Einstein gravity can be read as a remarkably strict sentence written in the language of geometry: take a Lorentzian metric g_{ab}, insist on a torsion-free, metric-compatible connection, build curvature, and then constrain that curvature by the Einstein equations. The payoff is that horizons, singularities, and causal structure become geometric facts rather than force-law add-ons. The price is that the theory is rigid, and rigidity invites two kinds of pressure.

The first is ultraviolet pressure. Even if one does not believe gravity is fundamentally described by a local metric field theory at arbitrarily short distances, one can still treat Einstein gravity as the leading term of an effective field theory. In that viewpoint, higher-curvature terms are not an aesthetic choice; they are the generic correction terms allowed by symmetry once you integrate out unknown microphysics. The second pressure is infrared and phenomenological: cosmic acceleration, galaxy-scale anomalies, and tensions between low-redshift structure growth and early-universe inferences motivate asking how far one can stretch geometry before it breaks consistency with the data.

Black holes are a particularly sharp testing ground. A black hole spacetime is not just strong field in the sense of large potentials, it is also a crucible for global concepts: event horizons are defined using causal structure that reaches to infinity, and curvature invariants probe geometric regularity in a coordinate-free way. If a modification changes causal cones, the very definition of horizon may change. If a modification adds extra degrees of freedom, the question becomes whether the horizon can hide them or whether they leak out as hair.

22.3 Minimal vocabulary: connections, torsion, and non-metricity

A differentiable manifold becomes a stage for physics only after we decide how to compare tensors at nearby points. That choice is encoded in a connection ∇, or equivalently, connection coefficients $\Gamma^a{}_{bc}$ in a chart. In standard general relativity, ∇ is the Levi-Civita connection of g_{ab}, defined uniquely by:

$$\nabla_c g_{ab} = 0 \quad \text{and} \quad T^a{}_{bc} \equiv 2\Gamma^a{}_{[bc]} = 0.$$

Modified geometric theories often relax one or both conditions.

Torsion is the antisymmetric part of the connection. In a torsionful geometry, parallelograms fail to close, and infinitesimal transport around a loop can depend on the loop's orientation in a way not captured by curvature alone. Torsion is naturally sourced by spin in Einstein-Cartan theory [3, 4].

Non-metricity measures failure of metric compatibility, $Q_{cab} \equiv -\nabla_c g_{ab}$. In metric-affine theories, both torsion and non-metricity can be present, and the metric and connection become independent variables.

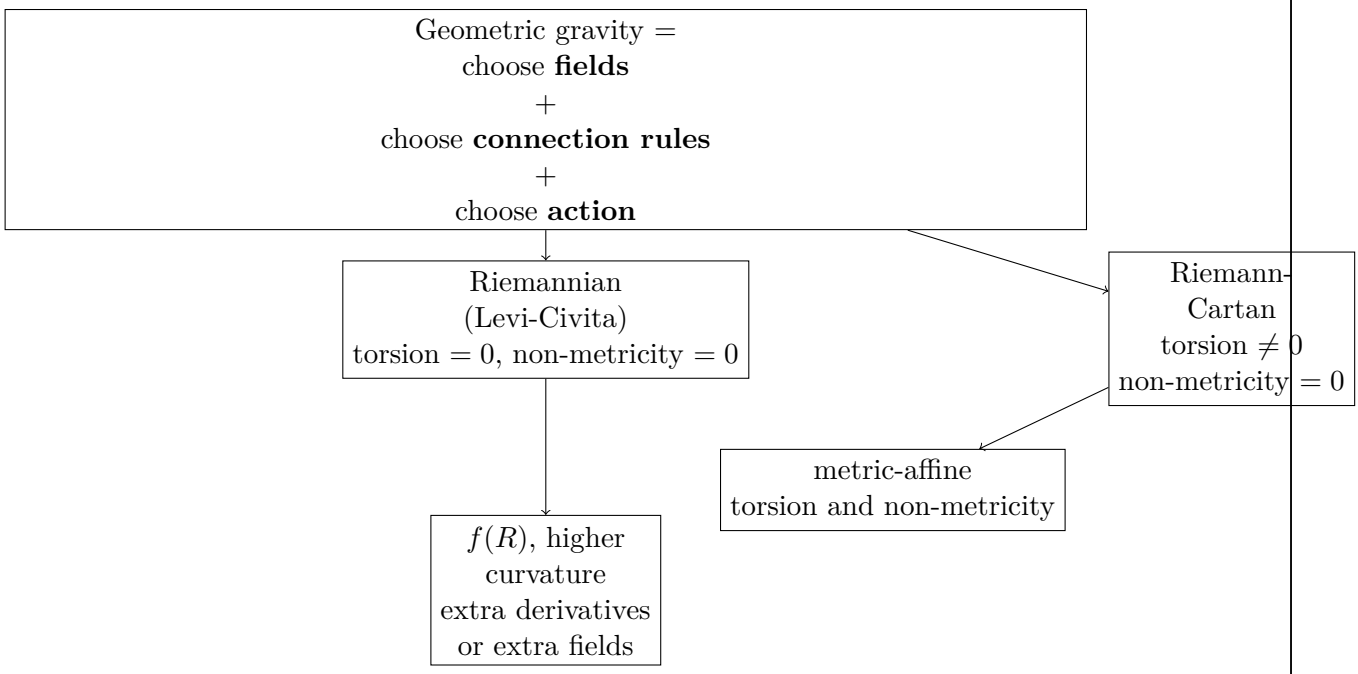


Figure 22.1: Conceptual map of geometric choices. This diagram is schematic and is included to orient the chapter, while the ten data figures later are strictly measurement-based.

22.4 $f(R)$ and higher-curvature terms: curvature becomes dynamical

22.4.1 From Einstein-Hilbert to $f(R)$

A prototypical extension replaces the Einstein-Hilbert action with a function of the Ricci scalar:

$$S = \frac{1}{16\pi G}\int d^4x\sqrt{-g}\, f(R) + S_{\rm matter}[g,\Psi].$$

In the metric formalism, varying with respect to g_{ab} yields fourth-order equations, schematically:

$$f_R R_{ab} - \frac{1}{2} f g_{ab} - \nabla_a\nabla_b f_R + g_{ab}\Box f_R = 8\pi G\, T_{ab},$$

where $f_R \equiv df/dR$. The most important structural fact is that f_R behaves like an extra scalar degree of freedom. Many $f(R)$ models are equivalent to scalar-tensor theories with a specific potential [1, 2].

22.4.2 Higher-curvature corrections as effective field theory

A broad, careful way to talk about higher curvature is to treat it as an effective theory:

$$S \approx \frac{1}{16\pi G}\int d^4x\sqrt{-g}\left(R + \alpha R^2 + \beta R_{ab}R^{ab} + \gamma R_{abcd}R^{abcd} + \cdots\right),$$

with coefficients suppressed by a high scale. In this framing, one does not claim that the higher-derivative equations are fundamental at all scales. One claims that for curvature below some cutoff, the corrections encode the leading influence of unknown ultraviolet physics.

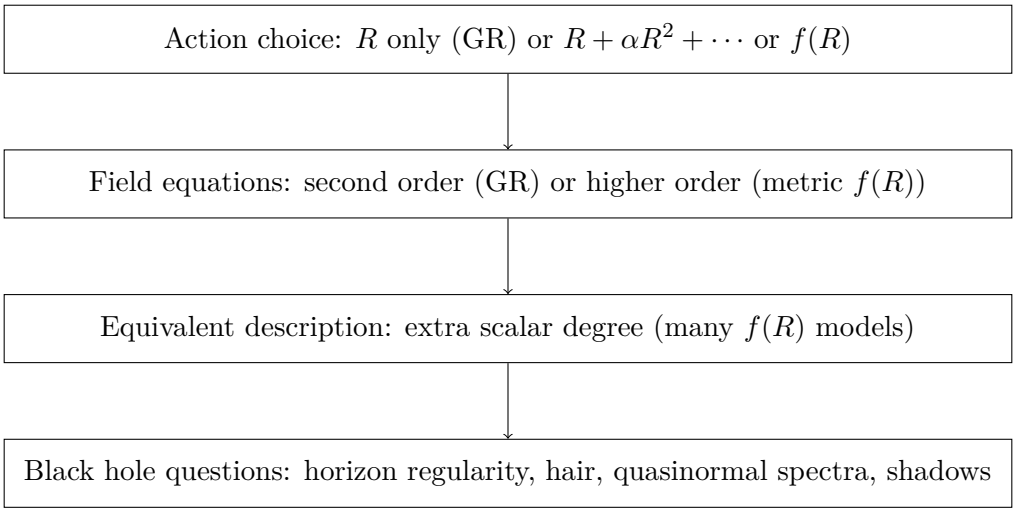

Action choice: R only (GR) or $R + \alpha R^2 + \cdots$ or $f(R)$
Field equations: second order (GR) or higher order (metric $f(R)$)
Equivalent description: extra scalar degree (many $f(R)$ models)
Black hole questions: horizon regularity, hair, quasinormal spectra, shadows

Figure 22.2: Conceptual pipeline from action modification to black hole observables. The plot is schematic; measurement-based constraints appear later.

22.5 Torsion theories: Einstein-Cartan as the clean starting point

Einstein-Cartan theory keeps the metric but allows torsion, with the connection determined by both the metric and the spin density of matter. In the simplest setup torsion is typically non-propagating: it is algebraically related to spin currents and does not introduce long-range torsion waves [3, 4].

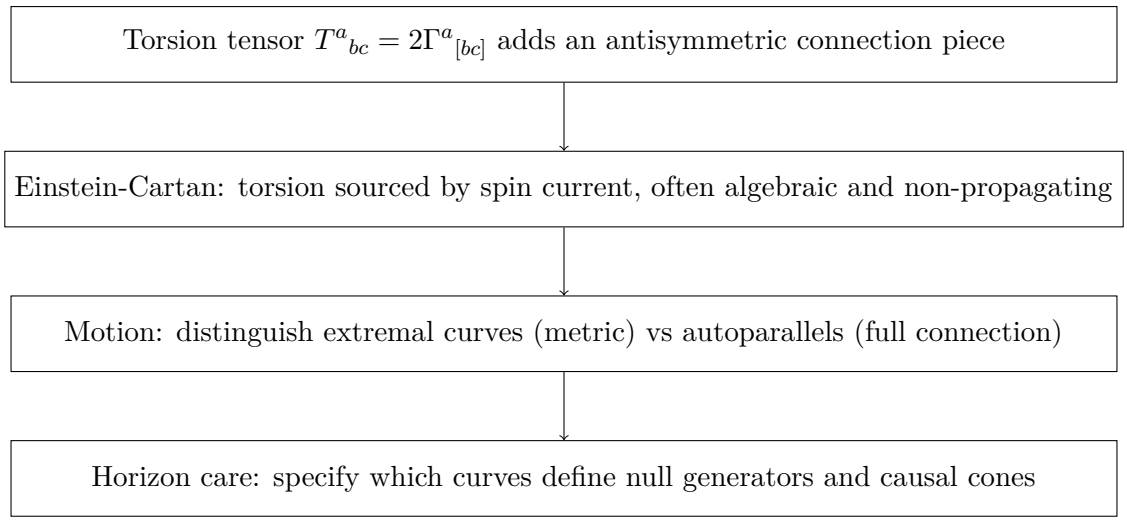

Torsion tensor $T^a{}_{bc} = 2\Gamma^a{}_{[bc]}$ adds an antisymmetric connection piece
Einstein-Cartan: torsion sourced by spin current, often algebraic and non-propagating
Motion: distinguish extremal curves (metric) vs autoparallels (full connection)
Horizon care: specify which curves define null generators and causal cones

Figure 22.3: Conceptual summary of torsion logic for black hole discussions.

22.6 Scalar-tensor gravity and black hole hair

Scalar-tensor theories add one or more scalar fields ϕ that couple to curvature or matter. No-hair theorems exist under specific assumptions [5, 6], and loopholes arise when those assumptions fail (curvature couplings, time-dependent scalars, nonminimal interactions).

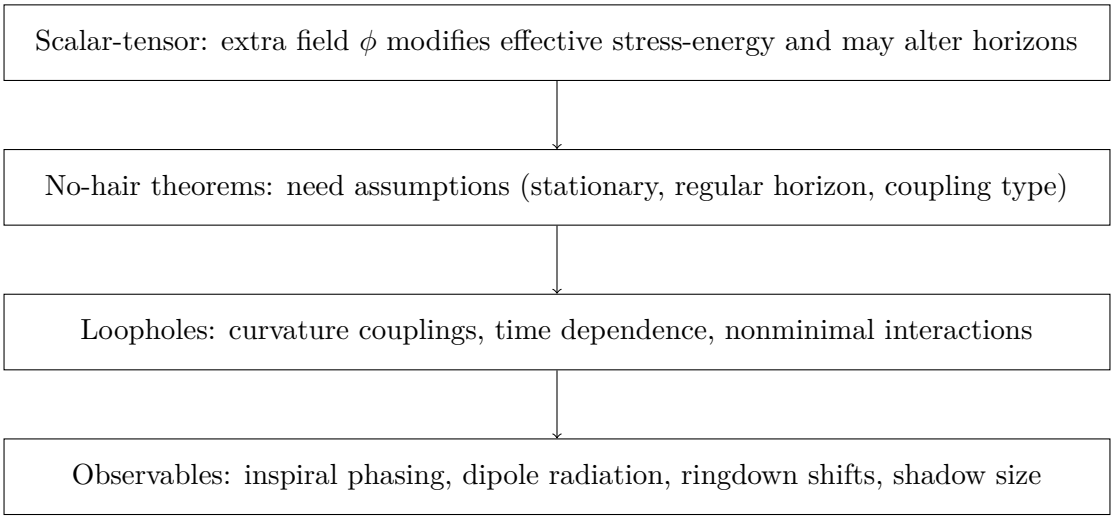

Scalar-tensor: extra field ϕ modifies effective stress-energy and may alter horizons
No-hair theorems: need assumptions (stationary, regular horizon, coupling type)
Loopholes: curvature couplings, time dependence, nonminimal interactions
Observables: inspiral phasing, dipole radiation, ringdown shifts, shadow size

Figure 22.4: Conceptual diagram of how scalar hair enters the black hole conversation.

22.7 Effective metrics, characteristic cones, and what observers actually measure

In extended theories, linearized excitations can propagate on effective metrics that differ from the matter metric. Multi-messenger timing constrains the relevant characteristic cones [12].

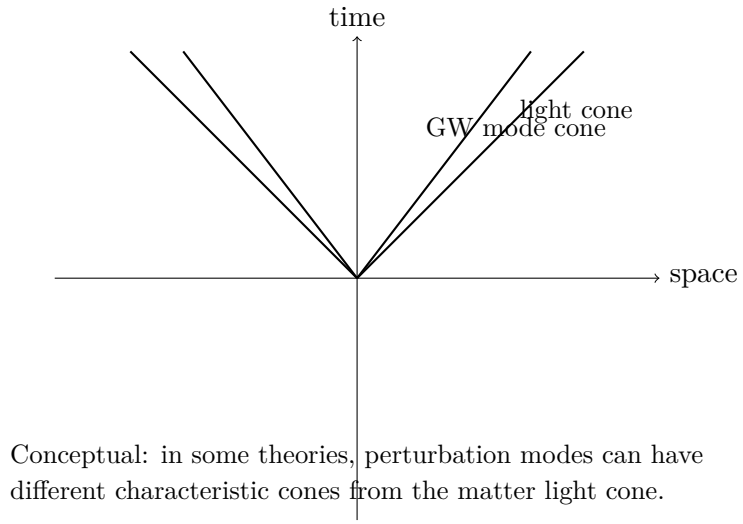


Figure 22.5: Conceptual cone picture. The data-based multi-messenger bound is plotted later in Fig. 22.8.

22.8 Data and analysis section: ten measurement-based figures

22.8.1 Solar System curvature per unit mass: constraints on PPN γ

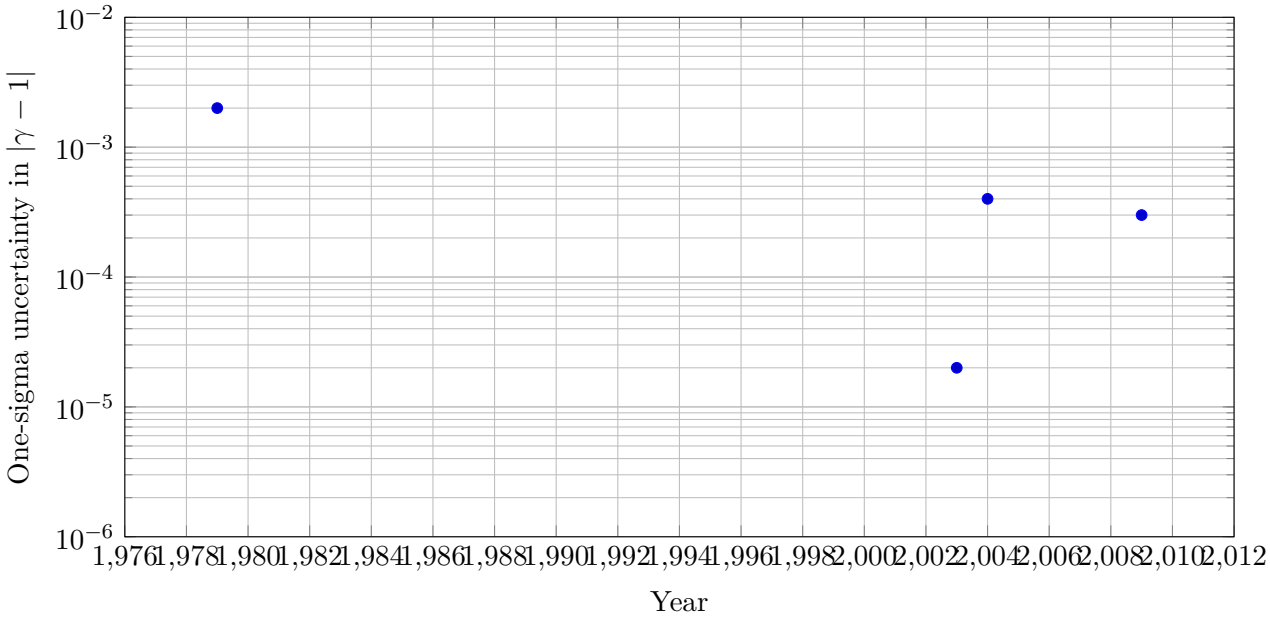


Figure 22.6: Measurement-based tightening of the PPN parameter γ. Points represent one-sigma uncertainties reported by Viking ranging (1979) [8], geodetic VLBI summary result referenced by VLBA analyses (2004) [9], a dedicated VLBA deflection measurement (2009) [9], and Cassini ranging (2003) [7].

22.8.2 Binary pulsar gravitational radiation: orbital decay ratios

22.8.3 Multi-messenger timing: bound on the speed of gravity

22.8.4 Graviton mass bounds from dispersion and ephemerides

22.8.5 Cosmological growth: $f\sigma_8(z)$ data with a reference curve

22.8.6 Horizon-scale imaging: ring diameters and inferred gravitational-radius angular sizes

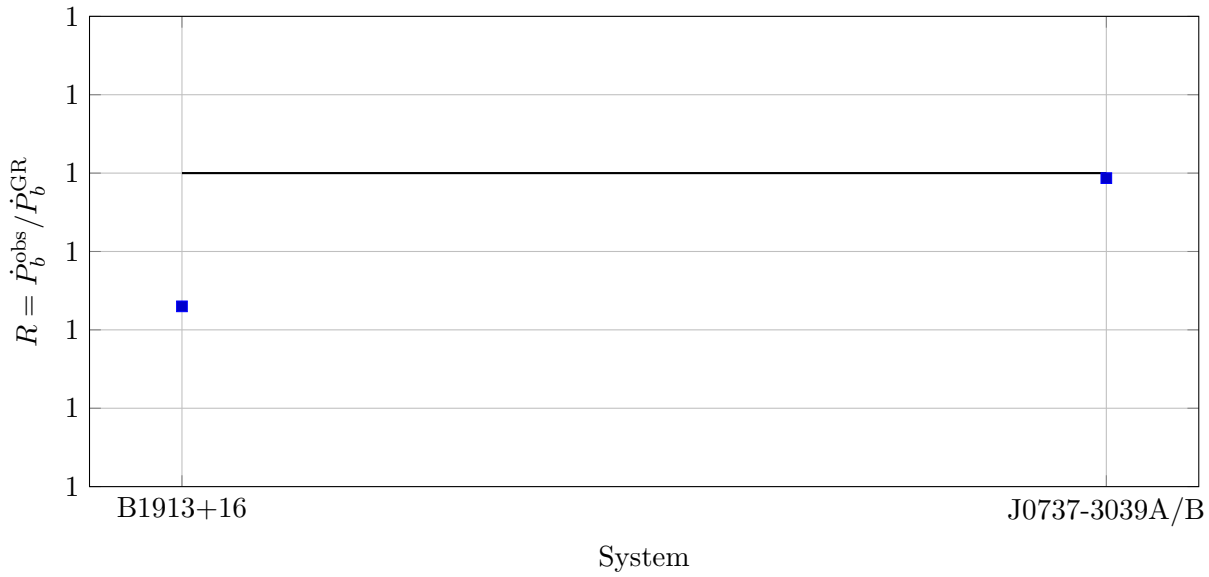


Figure 22.7: Measurement-based strong-field tests using binary pulsar orbital decay. The Hulse-Taylor ratio is reported as 0.9983 ± 0.0016 [10]. The Double Pulsar ratio is reported as 0.999936 with uncertainty at the 10^{-5} level [11].

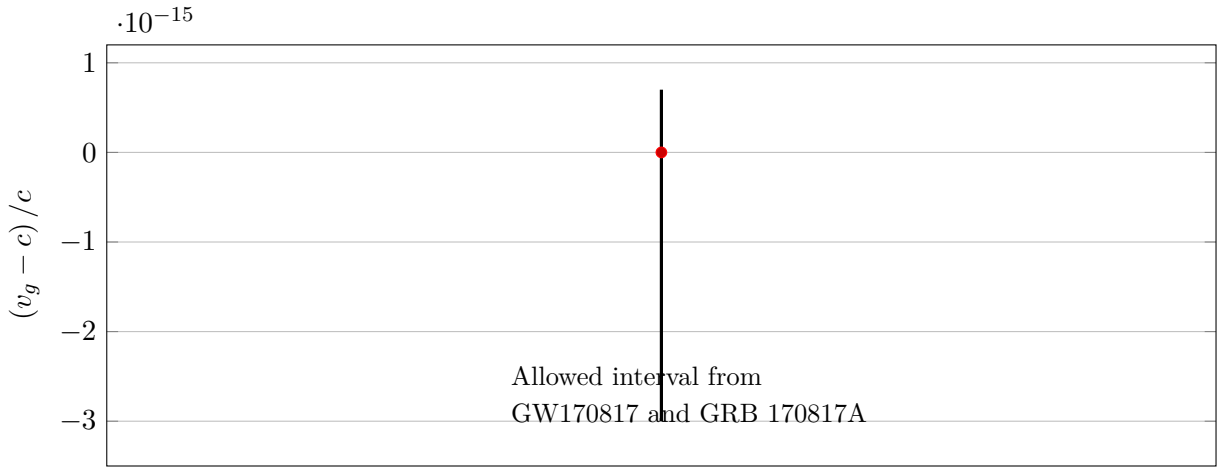


Figure 22.8: Measurement-based bound on the propagation speed of the observed gravitational-wave mode from GW170817 and GRB 170817A: $(v_g - c)/c$ lies between -3×10^{-15} and $+7 \times 10^{-16}$ [12].

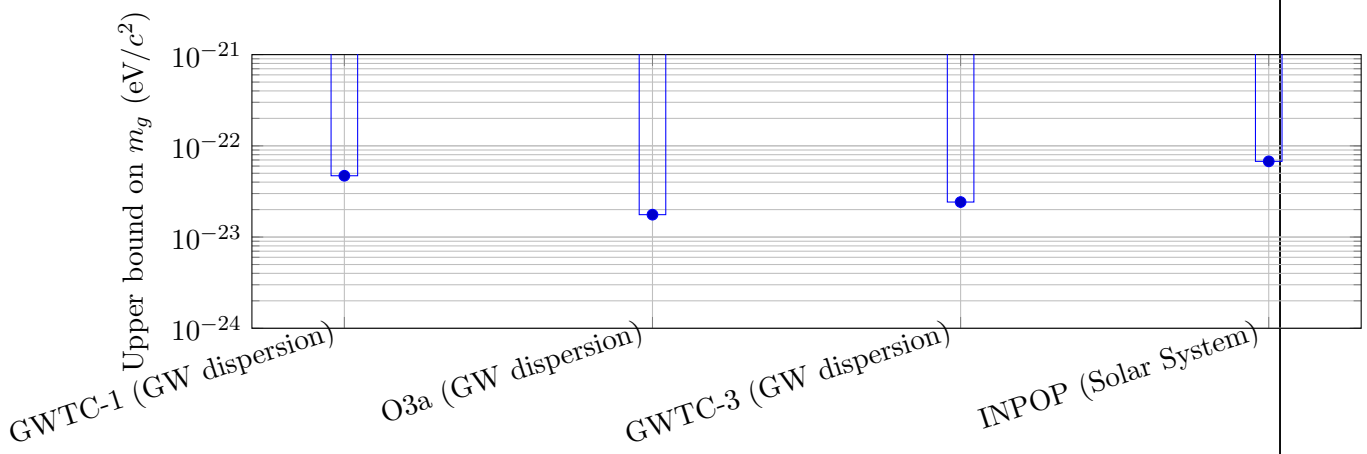


Figure 22.9: Measurement-based compilation of upper bounds on the graviton mass. The GWTC-1 combined bound $m_g \leq 4.7 \times 10^{-23}$ eV/c^2 is reported in tests with GWTC-1 [13]. An O3a-era bound $m_g \leq 1.76 \times 10^{-23}$ eV/c^2 is reported in LIGO documentation [15]. The GWTC-3 bound $m_g \leq 2.42 \times 10^{-23}$ eV/c^2 is reported in tests with GWTC-3 [14]. A Solar System ephemeris bound $m_g \leq 6.76 \times 10^{-23}$ eV/c^2 is reported using INPOP residual degradation [16].

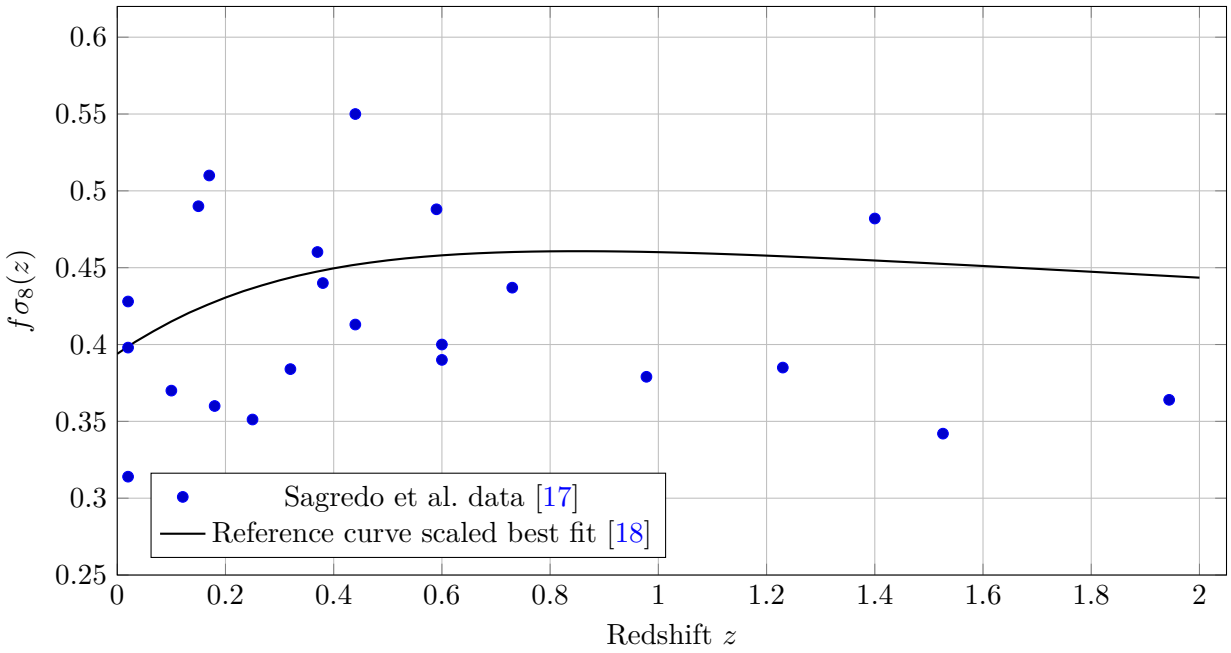


Figure 22.10: Measurement-based $f\sigma_8(z)$ compilation from Sagredo et al. [17] with a reference growth curve computed using representative Planck 2018 parameters [18] and scaled by a best-fit factor.

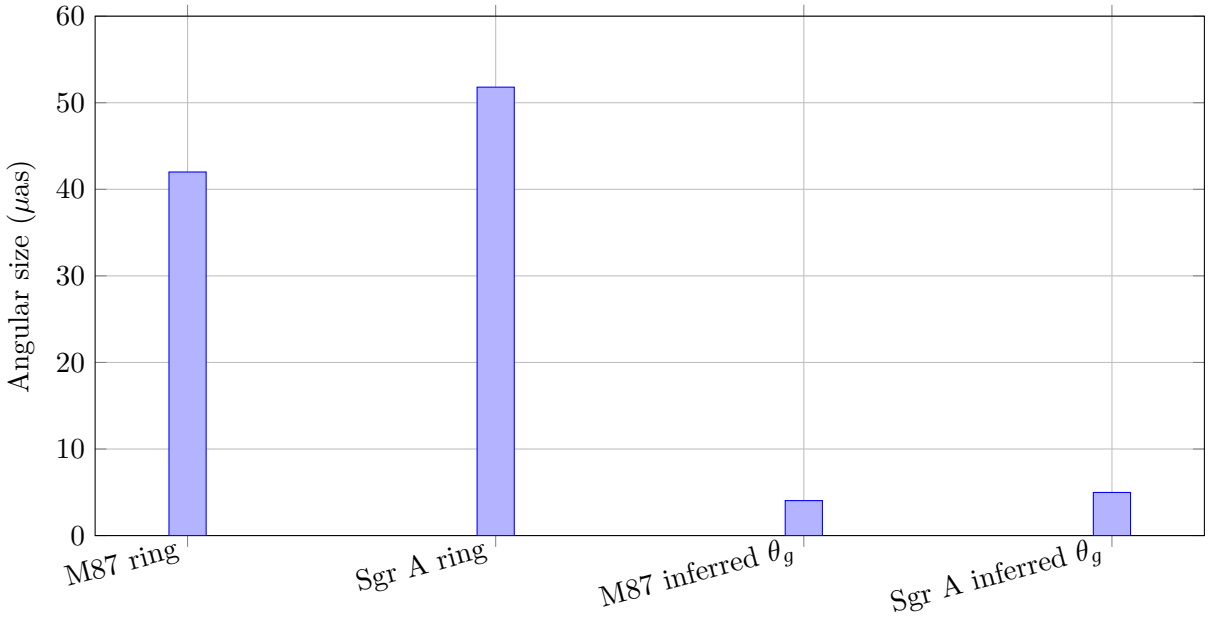


Figure 22.11: Measurement-based horizon-scale ring diameters and an algebraic inference of θ_g. The ring diameter for M87 is taken from EHT results [19]. The ring diameter for Sgr A is taken from EHT geometric modeling results [20]. Inferred θ_g values use $\theta_g = \theta_{\mathrm{ring}}/(2\sqrt{27})$ and inherit the measurement uncertainties.

22.9 Conclusion

Modified gravity is often presented as a list of Lagrangians, but the deeper unifying thread is geometric: which structures are fundamental, which are derived, and which define causality. In $f(R)$ and higher-curvature models, curvature itself becomes a carrier of additional dynamics, typically equivalent to new fields that reshape the meaning of vacuum. In torsion theories, the connection gains new algebraic content that can couple to spin and subtly modify the distinction between extremal curves and parallel-transported curves. In scalar-tensor models, the field content expands and the horizon becomes a boundary that either suppresses the extra structure or advertises it through hair.

Across the observational landscape, geometry is disciplined by measurements that span Solar System ranging, pulsar timing, multi-messenger timing, gravitational-wave dispersion tests, horizon-scale imaging, and cosmological growth. Many modifications can be written down, but very few survive the combined cross-scale constraints without screening, tuning, or a careful effective interpretation. That is not a failure. It is an empirical statement about how tightly spacetime geometry is already pinned down.

Bibliography

[1] T. P. Sotiriou and V. Faraoni, "$f(R)$ theories of gravity," *Rev. Mod. Phys.*, vol. 82, no. 1, pp. 451–497, 2010. [Online]. Available: https://doi.org/10.1103/RevModPhys.82.451

[2] A. De Felice and S. Tsujikawa, "$f(R)$ theories," *Living Rev. Relativity*, vol. 13, no. 3, 2010. [Online]. Available: https://doi.org/10.12942/lrr-2010-3

[3] F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, "General relativity with spin and torsion: foundations and prospects," *Rev. Mod. Phys.*, vol. 48, no. 3, pp. 393–416, 1976. [Online]. Available: https://doi.org/10.1103/RevModPhys.48.393

[4] I. L. Shapiro, "Physical aspects of the space-time torsion," *Phys. Rep.*, vol. 357, no. 2, pp. 113–213, 2002. [Online]. Available: https://arxiv.org/abs/hep-th/0103093

[5] S. W. Hawking, "Black holes in the Brans-Dicke theory of gravitation," *Commun. Math. Phys.*, vol. 25, pp. 167–171, 1972. [Online]. Available: https://projecteuclid.org/journals/communications-in-mathematical-physics/volume-25/issue-2/Black-holes-in-the-Brans-Dicke-theory-of-gravitation/cmp/1103857885.pdf

[6] J. D. Bekenstein, "Novel no-scalar-hair theorem for black holes," *Phys. Rev. D*, vol. 51, no. 12, pp. 6608–6611, 1995. [Online]. Available: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.51.R6608

[7] B. Bertotti, L. Iess, and P. Tortora, "A test of general relativity using radio links with the Cassini spacecraft," *Nature*, vol. 425, pp. 374–376, 2003. [Online]. Available: https://doi.org/10.1038/nature01997

[8] R. D. Reasenberg *et al.*, "Viking relativity experiment: verification of signal retardation by solar gravity," *Astrophys. J. Lett.*, vol. 234, pp. L219–L221, 1979. [Online]. Available: https://articles.adsabs.harvard.edu/pdf/1979ApJ...234L.219R

[9] E. Fomalont, S. Koopmans, and J. A. Benson, "Progress in measurements of the gravitational bending of radio waves using the VLBA," 2009. [Online]. Available: https://arxiv.org/pdf/0904.3992.pdf

[10] J. M. Weisberg and Y. Huang, "Relativistic measurements from timing the binary pulsar PSR B1913+16," *Astrophys. J.*, vol. 829, no. 1, 2016. [Online]. Available: https://arxiv.org/abs/1606.02744

[11] M. Kramer *et al.*, "Strong-field gravity tests with the Double Pulsar," *Phys. Rev. X*, vol. 11, 041050, 2021. [Online]. Available: https://arxiv.org/abs/2112.06795

[12] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A," *Astrophys. J. Lett.*, vol. 848, L13, 2017. [Online]. Available: https://arxiv.org/abs/1710.05834

[13] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Tests of general relativity with the binary black hole signals from the LIGO-Virgo Catalog GWTC-1," *Phys. Rev. D*, vol. 100, 104036, 2019. [Online]. Available: https://arxiv.org/abs/1903.04467

[14] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Tests of general relativity with GWTC-3," 2021. [Online]. Available: https://arxiv.org/abs/2112.06861

[15] LIGO Scientific Collaboration and Virgo Collaboration, "Tests of General Relativity with binary black holes from the second half of the third observing run (O3a)," PDF documentation, 2023. [Online]. Available: https://ligo.org/wp-content/uploads/2023/10/o3a_tgr.pdf

[16] A. Hees *et al.*, "Constraining the mass of the graviton with the planetary ephemeris INPOP," 2019. [Online]. Available: https://arxiv.org/pdf/1901.04307

[17] B. Sagredo, S. Nesseris, and D. Sapone, "Internal robustness of growth rate data," *Phys. Rev. D*, vol. 98, 083543, 2018. [Online]. Available: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.98.083543

[18] N. Aghanim *et al.* (Planck Collaboration), "Planck 2018 results. VI. Cosmological parameters," *Astron. Astrophys.*, vol. 641, A6, 2020. [Online]. Available: https://arxiv.org/abs/1807.06209

[19] Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope Results. IV. Imaging the central supermassive black hole," *Astrophys. J. Lett.*, vol. 875, L4, 2019. [Online]. Available: https://arxiv.org/abs/1906.11241

[20] Event Horizon Telescope Collaboration, "First Sagittarius A* Event Horizon Telescope Results. IV. Variability, morphology, and black hole mass," 2023. [Online]. Available: https://arxiv.org/abs/2311.08697

Black Holes Geometry: Curvature and Horizons

Chapter 23: Numerical Relativity as Computational Differential Geometry

Contents

Abstract ii

Keywords ii

0.1 Introduction: why numerical relativity is geometry with a compiler . . . ii

0.2 Discretizing geometry: from smooth manifolds to computable operators . iii

0.2.1 The 3+1 split as a computational chart iii

0.2.2 Meshes, patches, and why "coordinates are a polite lie" becomes operational iv

0.2.3 Finite differences, finite elements, and spectral methods as discrete connections v

0.3 Constraint damping and geometric stability vi

0.3.1 Why constraints are an active dynamical system vi

0.3.2 BSSN and moving punctures vi

0.3.3 Generalized harmonic gauge and constraint damping vii

0.4 Extracting gravitational waves: curvature at infinity vii

0.4.1 From local curvature to radiative degrees of freedom vii

0.4.2 Finite-radius extraction, extrapolation, and Cauchy-characteristic extraction viii

0.5 Horizon tracking and quasi-local invariants viii

0.5.1 Apparent horizons as marginally trapped surfaces viii

0.5.2 Mass, spin, and multipoles from a surface ix

0.6 Data-driven illustrations from open gravitational-wave catalogs x

0.6.1 Dataset provenance and derived quantities x

0.6.2 Ten dataset-based figures and what they imply for numerical geometry xii

0.7 A workflow view: what a modern NR pipeline is really doing xvi

0.8 Phenomenon statements: real-world applicability of computational differential geometry xvi

0.9 Conclusion xx

Abstract

Numerical relativity is often introduced as "solving Einstein's equations on a computer". That slogan is correct but incomplete. What is actually being discretized is not merely a system of partial differential equations, but a whole geometric world: a Lorentzian manifold decomposed into spacelike slices, a connection encoded through Christoffel symbols and spin coefficients, curvature reconstructed from finite or spectral approximations, and global invariants inferred from quasi-local surfaces that move through the computational domain. In this chapter we treat numerical relativity as computational differential geometry: a disciplined translation of smooth curvature into stable numerical operators, with error control that respects constraints and causal structure. We connect core algorithmic choices (finite differences, finite elements, spectral methods, gauge drivers, constraint damping, wave extraction, horizon finding, and geometric diagnostics) to observational consequences by building ten data-driven illustrations from the open gravitational-wave event catalog GWTC-3 via the Gravitational Wave Open Science Center (GWOSC) Event Portal and API [1–4]. In addition, six conceptual diagrams summarize the geometric flow of a modern simulation and analysis pipeline. The goal is a practical geometric intuition: the reader should be able to look at a numerical output, ask which geometric object it approximates, identify which constraints protect it, and understand which diagnostics certify it.

Keywords

Keywords: numerical relativity; 3+1 decomposition; BSSN; generalized harmonic gauge; constraint damping; adaptive mesh refinement; spectral methods; Weyl scalar Ψ_4; Bondi news; Cauchy-characteristic extraction; apparent horizons; quasi-local invariants; gravitational waves; GWTC-3; GWOSC.

0.1 Introduction: why numerical relativity is geometry with a compiler

Numerical relativity is the place where smooth Lorentzian geometry meets finite memory, finite time steps, and the awkward truth that a coordinate system is both a gauge choice and a numerical instrument. A simulation of two black holes is not just a computed metric g_{ab}, it is a controlled approximation to a geometric equivalence class, and the

difference matters. Two coordinate representations can encode the same spacetime, yet one may be numerically stable while the other can tear itself apart by amplifying constraint-violating modes. The practical question is therefore geometric: which formulation of the Einstein equations treats the constraints as guardians rather than decorations, and which gauge choices make the physically interesting hypersurfaces accessible without generating numerical pathologies.

A productive way to think about this chapter is to imagine that you are building a laboratory for curvature. The laboratory contains: (i) an initial data set that satisfies the Hamiltonian and momentum constraints; (ii) an evolution rule that preserves those constraints in the continuum and damps violations in the discrete; (iii) a method to compute radiation at large distances, ideally at future null infinity; and (iv) quasi-local tools to locate horizons and measure mass, spin, and multipole moments. Each item is a geometric operator dressed in computational clothing. The computations become meaningful when they map to invariants or controlled gauge-dependent proxies. For gravitational-wave astronomy, these proxies are not optional: they flow directly into waveform models, parameter estimation, and precision tests of general relativity [2, 4].

This chapter is organized around five geometric tasks that every large-scale numerical relativity code must accomplish: discretize geometry, stabilize constraints, extract radiation, track horizons, and diagnose curvature. We end by grounding these tasks in open event-catalog data. The catalog figures are not "simulations of a simulation"; they are derived from publicly released measurements of real detections, and they are used here to clarify which regions of parameter space and which physical scales push numerical methods the hardest. All ten data-driven figures in Section 0.6 are dataset-based. The six conceptual diagrams are explicitly labeled as conceptual aids.

0.2 Discretizing geometry: from smooth manifolds to computable operators

0.2.1 The 3+1 split as a computational chart

The standard entry point to numerical relativity is the 3+1 (ADM) decomposition . One foliates spacetime by a family of spacelike hypersurfaces Σ_t labeled by a time parameter t. The line element is written as

$$ds^2 = -\alpha^2 dt^2 + \gamma_{ij}\Big(dx^i + \beta^i dt\Big)\Big(dx^j + \beta^j dt\Big), \tag{1}$$

where α is the lapse function, β^i is the shift vector, and γ_{ij} is the induced spatial metric on Σ_t. This is not merely a decomposition; it is a commitment to a computational strategy. The lapse and shift determine how the numerical grid threads through spacetime, and therefore how coordinate singularities, slice stretching, and near-horizon behavior manifest

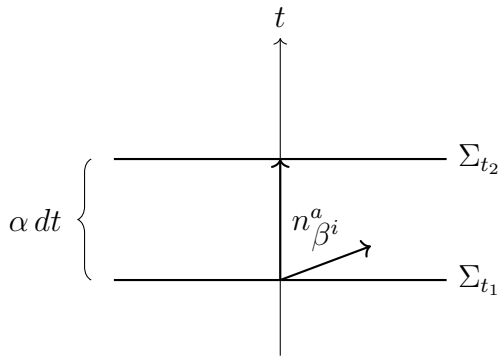


Figure 1: Conceptual Diagram 23.A: A 3+1 foliation with lapse α, shift β^i, and unit normal n^a. The choice of (α, β^i) is both gauge freedom and a numerical steering wheel.

on a mesh.

The extrinsic curvature captures how each hypersurface is embedded in the full spacetime:

$$K_{ij} = -\frac{1}{2\alpha}\Big(\partial_t \gamma_{ij} - D_i\beta_j - D_j\beta_i\Big), \tag{2}$$

with D_i the covariant derivative compatible with γ_{ij}. The Einstein equations split into constraints on each slice and evolution equations that propagate (γ_{ij}, K_{ij}) in time. The constraints are

$$\mathcal{H} := {}^{(3)}R + K^2 - K_{ij}K^{ij} - 16\pi\rho = 0, \tag{3}$$

$$\mathcal{M}^i := D_j\Big(K^{ij} - \gamma^{ij}K\Big) - 8\pi S^i = 0, \tag{4}$$

where ${}^{(3)}R$ is the Ricci scalar of γ_{ij}, and (ρ, S^i) are projections of the stress-energy tensor. In vacuum, $\rho = S^i = 0$. The constraints are not optional bookkeeping. Numerically, they are the difference between a simulation that stays on the manifold of solutions and a simulation that drifts into an unphysical space that only resembles a spacetime.

0.2.2 Meshes, patches, and why "coordinates are a polite lie" becomes operational

In differential geometry, we cover a manifold with coordinate charts. In numerical relativity we often cover the computational domain with coordinate patches, refinement boxes, or spectral subdomains. The similarity is not metaphorical. Overlap regions require interpolation (a discrete version of transition functions), and the numerical stability of the full code depends on whether these discrete transitions preserve constraints and respect causal propagation.

A subtle point: the computational "grid" is not a geometric object. The geometric object is the metric and its derivatives. The grid is a sampling strategy for representing

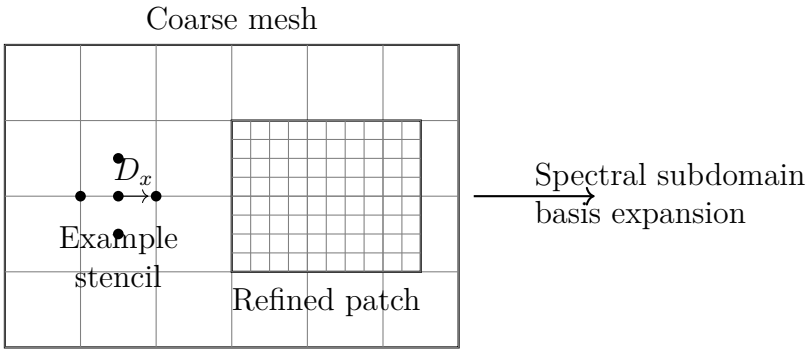


Figure 2: Conceptual Diagram 23.B: Mesh refinement and local stencils. The geometric goal is not "more points" but a controlled approximation of covariant derivatives and curvature in regions where gradients sharpen.

those objects. When the curvature grows, the grid must respond by refining, changing coordinates, or changing basis functions. Adaptive mesh refinement can be viewed as a dynamic choice of local chart resolution. Spectral domain decomposition can be viewed as a choice of multiple charts with global basis functions on each chart.

0.2.3 Finite differences, finite elements, and spectral methods as discrete connections

At the continuum level, the Levi-Civita connection provides a covariant derivative ∇_a that is metric-compatible and torsion-free. In a numerical method, one replaces derivatives by discrete operators. The geometric risk is that a naive discrete derivative does not respect product rules or metric identities, and therefore breaks the algebra that makes constraints propagate benignly.

Finite differences approximate derivatives by local stencils. Their appeal is simplicity and locality. Their risk is dispersion, dissipation, and instability near steep gradients unless carefully controlled. They become geometry-aware when used with summation-by-parts operators, constraint-preserving boundary conditions, and compatible interpolation across refinement boundaries.

Finite elements build an approximation space from basis functions on elements. From a geometric perspective, they become powerful when paired with ideas from discrete exterior calculus: differential forms are represented in compatible finite element spaces so that discrete analogs of $d^2 = 0$ and Stokes' theorem hold exactly. This is one of the cleanest ways to keep a geometric identity true after discretization, rather than hoping it becomes approximately true by accident.

Spectral and pseudo-spectral methods represent fields by global basis functions on subdomains. For smooth solutions they deliver exponential convergence, which is a geometric luxury: curvature becomes accurately computable at far fewer degrees of freedom. Their drawback is handling shocks and discontinuities (for example matter

interfaces in neutron star simulations) and managing complex domain decompositions.

In practice, numerical relativity uses hybrids: finite differences with adaptive mesh refinement for moving puncture black holes, and pseudo-spectral methods for excision-based black holes. The choice is partly about physics, but it is also about which discrete connection lets the constraints behave as a stable subsystem.

0.3 Constraint damping and geometric stability

0.3.1 Why constraints are an active dynamical system

Constraints are equalities that define a submanifold of the full state space of variables. In the continuum, if the evolution equations are consistent, constraint satisfaction is preserved. In discrete form, truncation error drives the system off that submanifold. If constraint violations grow, they feed back into the evolution equations and deform the simulated geometry.

A modern code therefore treats constraints as a dynamical system that must be controlled. The two main strategies are: (i) choose an evolution formulation whose constraint subsystem is well-posed and propagates violations away; and (ii) add damping terms that force violations to decay.

0.3.2 BSSN and moving punctures

The BSSN formulation begins by rewriting the ADM variables in a conformal decomposition that separates scale from shape. One defines a conformal metric $\tilde{\gamma}_{ij}$ with unit determinant, a conformal factor (often through ϕ or χ), and a trace-free conformal extrinsic curvature. Auxiliary variables such as the conformal connection functions $\tilde{\Gamma}^i$ are evolved. This rearrangement is not cosmetic. It changes the hyperbolic character of the system and improves numerical stability for strong-field dynamics [6, 7].

The moving puncture method is a complementary idea: represent black holes as punctures with suitable gauge conditions (commonly variants of 1+log slicing for the lapse and Gamma-driver conditions for the shift). Geometrically, the puncture is not a physical point of the manifold; it is a coordinate and representation choice that avoids explicit excision while still capturing horizon dynamics. In many practical simulations, the stability of moving punctures is a triumph of gauge as geometry-control, not just gauge as coordinate choice.

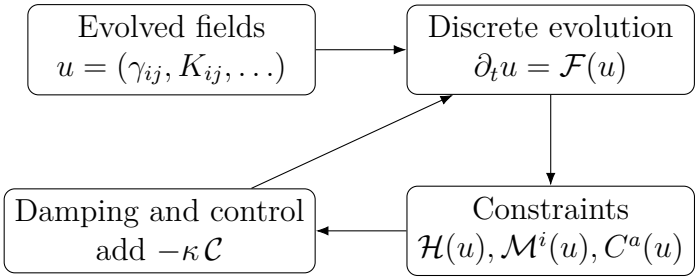


Figure 3: Conceptual Diagram 23.C: Constraint monitoring and damping as a feedback loop. Numerically, constraints are not only checked, they are controlled.

0.3.3 Generalized harmonic gauge and constraint damping

Generalized harmonic formulations promote coordinates to dynamical objects via gauge source functions H^a such that

$$\Box x^a = -H^a, \tag{5}$$

and define gauge constraints $C^a := H^a + \Gamma^a$, where $\Gamma^a = g^{bc}\Gamma^a_{bc}$. The evolution equations can be cast as a set of wave equations for the metric. The breakthrough in early successful binary black hole evolutions was not only the formulation itself, but the realization that constraint damping terms can stabilize the evolution by driving $C^a \to 0$ [8–10].

One can interpret constraint damping as adding a geometric "restoring force" toward the constraint surface. Importantly, the damping must preserve hyperbolicity and avoid introducing superluminal propagation of constraint effects. This is a case where computational choices are constrained by causal geometry.

0.4 Extracting gravitational waves: curvature at infinity

0.4.1 From local curvature to radiative degrees of freedom

In numerical relativity the gravitational wave is not a separate field. It is encoded in the geometry itself. The challenge is therefore to extract the radiative content of curvature in a way that is gauge robust and physically meaningful for distant observers.

A common approach uses the Newman-Penrose scalar Ψ_4 , defined (in one standard convention) as a component of the Weyl tensor C_{abcd} projected onto a null tetrad:

$$\Psi_4 = -C_{abcd} n^a \bar{m}^b n^c \bar{m}^d, \tag{6}$$

where n^a is an outgoing null vector and m^a is a complex null vector tangent to the sphere. This formalism traces back to the classic work of Newman and Penrose [13]. In

an asymptotically flat region and an appropriate tetrad choice, Ψ_4 relates to the second time derivative of the strain:

$$\Psi_4 \approx \ddot{h}_+ - i\,\ddot{h}_\times. \tag{7}$$

In practice the strain is obtained by time integration with care to remove drifts, and systematic uncertainties must be tracked.

A striking modern development is the explicit use of Ψ_4 in parameter inference pipelines, connecting the curvature-based output of numerical relativity more directly to data analysis [16]. This is a conceptual closure: curvature is computed, curvature is transported outward, and curvature is compared to detectors.

0.4.2 Finite-radius extraction, extrapolation, and Cauchy-characteristic extraction

Many codes compute waves on coordinate spheres at finite radius. One then extrapolates to infinity, attempting to remove gauge effects and finite-radius errors. This works well in many situations, but it is not guaranteed to be unambiguous.

Cauchy-characteristic extraction (CCE) provides a geometric alternative: one uses a standard 3+1 Cauchy evolution in the near zone and passes data on a worldtube to a characteristic evolution that naturally reaches future null infinity I^+, where radiation is defined unambiguously in the Bondi framework [14, 15]. The computational viewpoint is patching: different coordinate systems and evolution schemes cover different geometric regions, and their interface must preserve the correct transport of curvature.

0.5 Horizon tracking and quasi-local invariants

0.5.1 Apparent horizons as marginally trapped surfaces

Numerical relativity needs black holes to be measurable objects, not just regions of high curvature. The most common quasi-local diagnostic is the apparent horizon, a marginally outer trapped surface on a spatial slice. If s^i is the outward unit normal to a closed 2-surface $S \subset \Sigma_t$, and n^a is the timelike unit normal to the slice, one forms outgoing and ingoing null normals $\ell^a = n^a + s^a$ and $k^a = n^a - s^a$ (up to normalization). The expansions $\theta_{(\ell)}$ and $\theta_{(k)}$ measure how null congruences orthogonal to S expand. An apparent horizon satisfies

$$\theta_{(\ell)} = 0, \qquad \theta_{(k)} < 0. \tag{8}$$

Finding such a surface is a geometric root-finding problem on each time slice. Practical algorithms typically represent the surface by spherical harmonics or other parameterizations and solve a nonlinear elliptic equation. Thornburg's review remains a standard reference for the geometry and numerics of horizon finding [12].

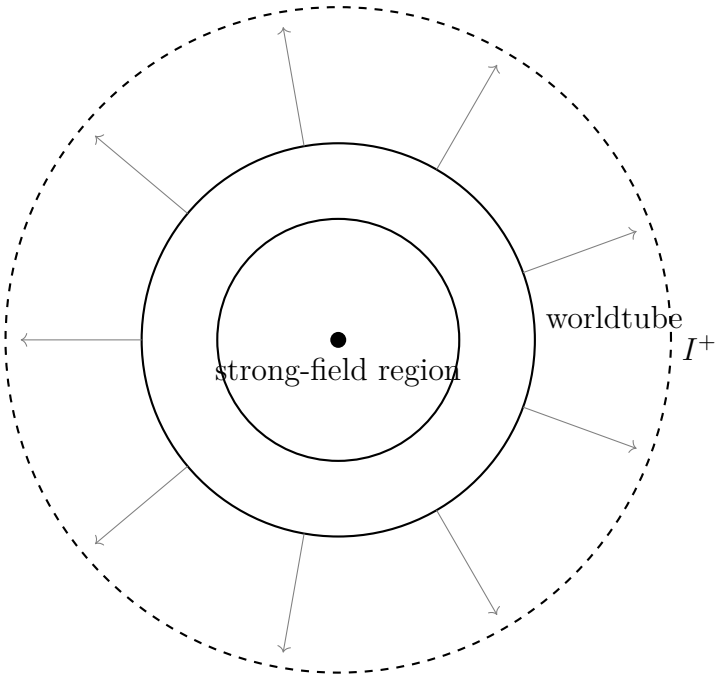


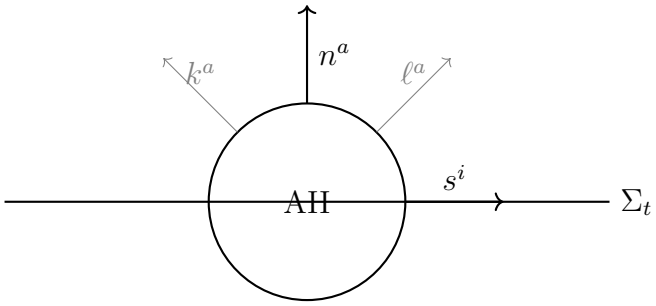
Figure 4: Conceptual Diagram 23.D: Wave extraction geometry. Finite-radius extraction uses spheres in the Cauchy region. CCE propagates the geometry to future null infinity, where Bondi news and strain are defined without finite-radius ambiguity [14, 15].

0.5.2 Mass, spin, and multipoles from a surface

Once an apparent horizon is located, it becomes a laboratory instrument. Its area A defines the irreducible mass

$$M_{\mathrm{irr}} = \sqrt{\frac{A}{16\pi}}. \tag{9}$$

Spin can be estimated by approximate Killing vectors on the horizon or through quasi-local angular momentum integrals. Multipole moments of the horizon's intrinsic curvature describe how far the horizon is from an exact Kerr geometry. These diagnostics are not only descriptive. They support convergence checks and quantify gauge effects. They also link simulations to observations by providing final mass and spin estimates used in ringdown modeling and tests of general relativity.



$\theta_{(\ell)} = 0$ defines a marginally trapped surface.

Figure 5: Conceptual Diagram 23.E: Apparent horizon (AH) as a marginally outer trapped surface on a slice. Horizon tracking converts "a black hole exists" into measured quasi-local invariants [12].

0.6 Data-driven illustrations from open gravitational-wave catalogs

0.6.1 Dataset provenance and derived quantities

This section uses open event-catalog data to make a concrete point: numerical relativity is forced to be accurate in precisely those regions of parameter space where the universe produces detections. We therefore build ten data-driven figures from the GWTC-3-confident set of events using the GWOSC Event Portal and its API [1–3]. The median parameter estimates used here are the portal's default values, and are suitable for illustrative geometry-aware analysis. A production-grade study would propagate full posterior samples.

From the catalog medians we form derived quantities that are directly connected to geometric and numerical questions:

- Mass ratio $q = m_1/m_2$. Large q stresses resolution requirements and waveform modeling.
- Detector-frame rescaling through $(1+z)$, for example $\mathcal{M}_{\rm det} = (1+z)\mathcal{M}$, reminding us that waveform phase evolution is tied to detector-frame masses.
- Radiated mass fraction $f_{\rm rad} = (M - M_f)/M$. This is a quasi-invariant statement about how much energy leaves the system as gravitational radiation, and it is a direct output of accurate strong-field evolution.
- An illustrative frequency scale, the Schwarzschild ISCO frequency:

$$f_{\rm ISCO} \approx \frac{c^3}{6^{3/2}\pi G\, M_{\rm det}} \approx \frac{4397\ {\rm Hz}}{M_{\rm det}/M_\odot}. \tag{10}$$

Table 1: Catalog fields used in the dataset-based figures, taken from GWOSC default parameter estimates [2, 3].

Field	**Meaning (used here)**
m_1, m_2	Source-frame component masses (median) in $M_\odot$.
$\mathcal{M}$	Source-frame chirp mass (median) in $M_\odot$.
$\chi_{\rm eff}$	Effective inspiral spin (median).
D_L	Luminosity distance (median) in Mpc.
z	Redshift (median).
SNR	Network matched-filter signal-to-noise ratio.
FAR	False alarm rate in yr^{-1}.
$p_{\rm astro}$	Astrophysical probability estimate.
M, M_f	Source-frame total and final masses (medians) in $M_\odot$.

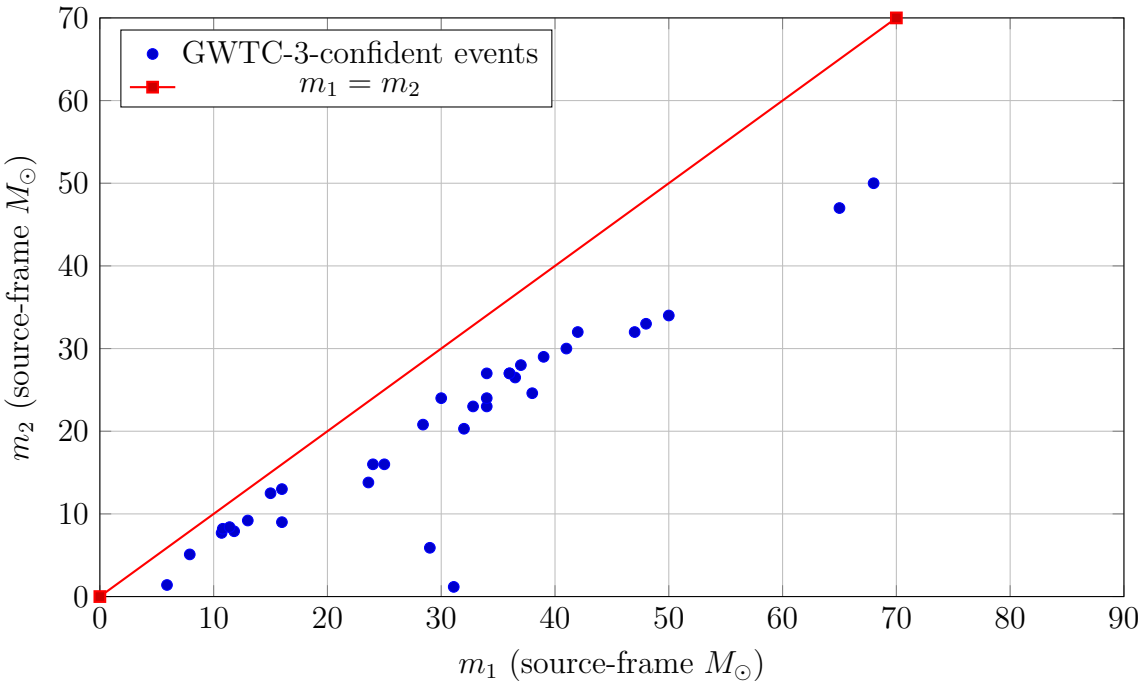


Figure 6: Data-driven Figure 23.1: Source-frame component masses for GWTC-3-confident events (35 detections) from GWOSC default parameter estimates [2–4]. The diagonal marks equal-mass binaries. This mass plane is the practical "target manifold" that waveform modeling and strong-field evolution must cover.

This mapping from mass to frequency is one reason compact binaries are audible in ground-based detectors.

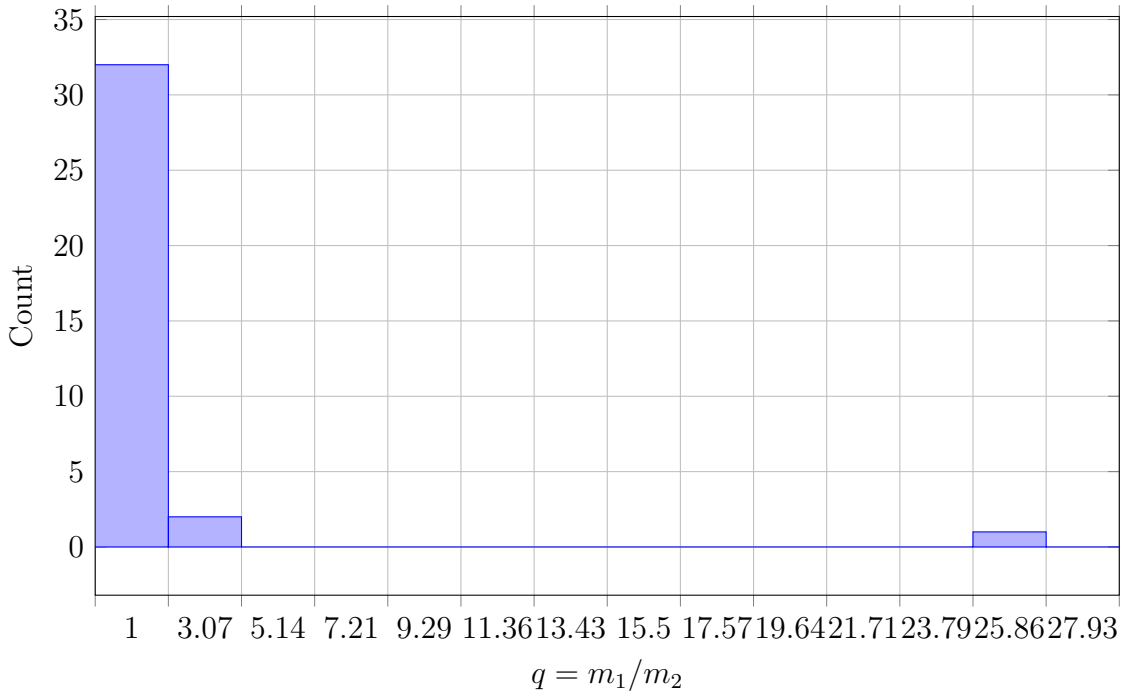


Figure 7: Data-driven Figure 23.2: Mass ratio distribution for GWTC-3-confident events from GWOSC default parameter estimates [3, 4]. A long tail toward high q indicates why stable methods and flexible discretizations matter, because unequal-mass systems generate multi-scale geometry.

0.6.2 Ten dataset-based figures and what they imply for numerical geometry

The mass plane is where computational differential geometry becomes operational. Near equal mass, symmetry helps, but the nonlinear merger still requires robust gauge and constraint control. At higher asymmetry (large q), resolution and numerical dissipation become more delicate. In both regimes, the geometry must be evolved long enough to produce waveforms with reliable phase.

The chirp mass is a geometric "clock" for inspiral: it controls how rapidly the orbital frequency sweeps upward. Numerically, that sweep is not just a feature of the waveform, it is a feature of how quickly the curvature changes, and therefore how small the time step must be to maintain stability and accuracy.

Spins are where numerical geometry and physical interpretation become tightly coupled. A modest change in spin can shift the location of the innermost stable orbit, alter precession, and change the waveform's harmonic content. These are geometric effects, and a code that cannot robustly evolve rotating spacetimes will not provide credible waveforms.

This figure is a reminder that numerical relativity does not live in isolation. The simulation is computed in geometric units tied to the source, but the inference problem is solved in the detector frame. The bridge is redshift and cosmology. Even if one is not doing

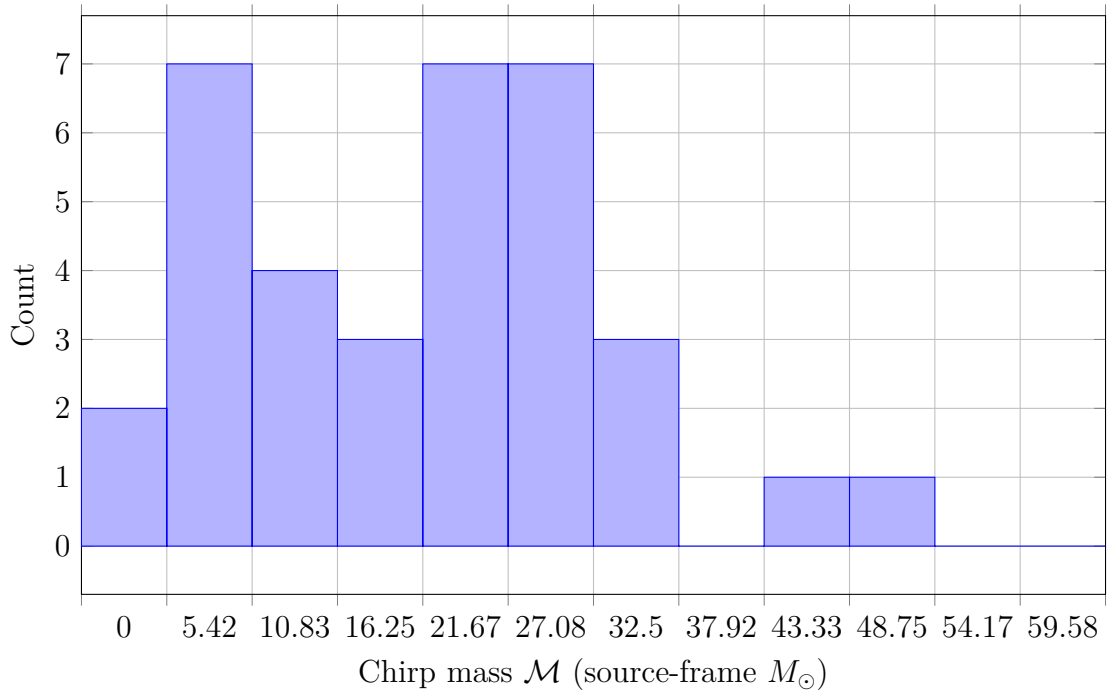


Figure 8: Data-driven Figure 23.3: Source-frame chirp mass distribution for GWTC-3-confident events [3, 4]. Chirp mass largely controls inspiral phasing, so this histogram can be read as a practical demand curve for phase-accurate numerical geometry.

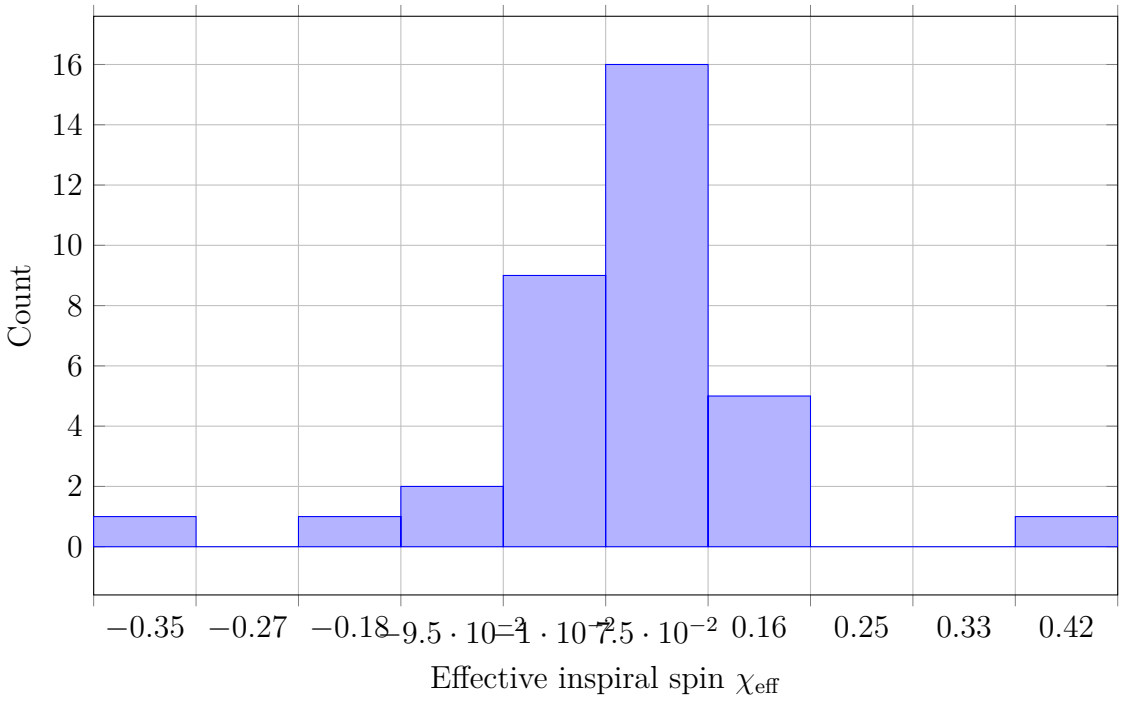


Figure 9: Data-driven Figure 23.4: Effective spin χ_{eff} for GWTC-3-confident events [3, 4]. Spin affects both strong-field geometry (frame dragging, horizon shape) and waveform morphology, increasing the importance of robust gauge choices and accurate curvature diagnostics.

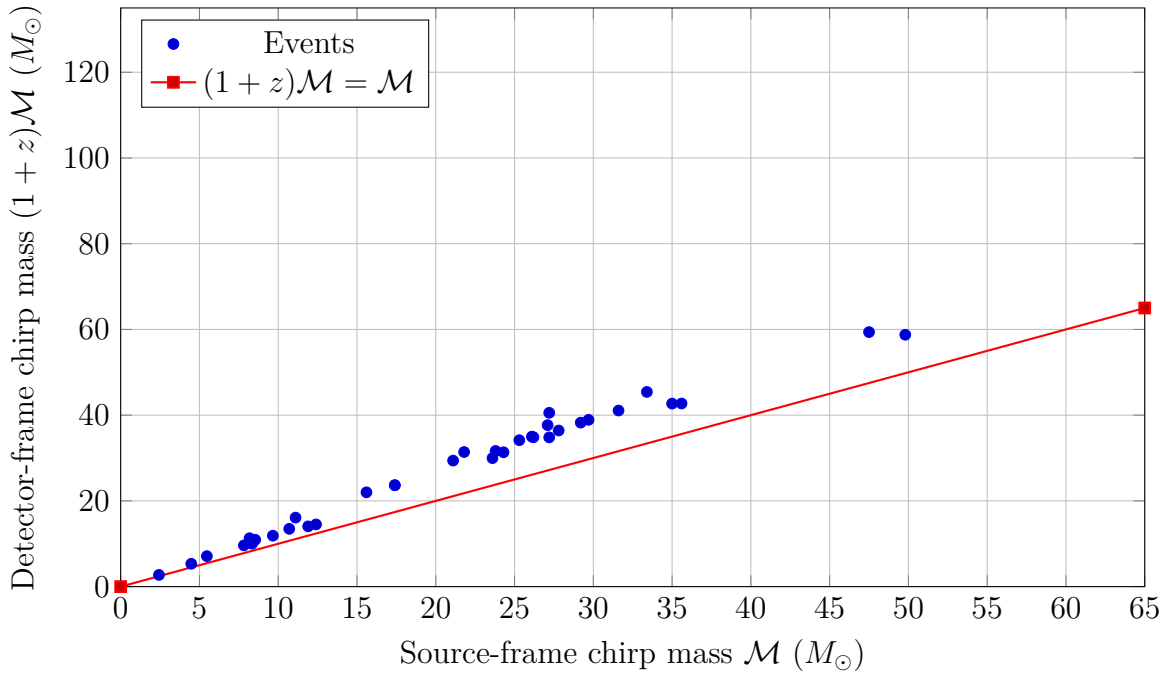


Figure 10: Data-driven Figure 23.5: Redshift rescaling from source-frame to detector-frame chirp mass using GWOSC default parameter estimates [3, 4]. This emphasizes a practical point: waveform phase evolution in the detector is governed by detector-frame masses, even when the source physics is described in its own frame.

cosmology, one is always doing redshift bookkeeping.

Radiated energy is not a perturbative detail. It is an integrated curvature effect. Numerically, it depends on stable constraint control, accurate wave extraction, and sufficient resolution to avoid artificial dissipation. A code can look stable yet systematically bias the radiated fraction if wave extraction is contaminated by gauge effects or finite-radius error.

The diagonal $M_f = M$ is forbidden in a real merger except in the trivial limit of no radiation. The separation from the diagonal is a global geometric effect. In computational terms it is a stress test of the full pipeline: evolution, gauge, extraction, and diagnostics.

A detection pipeline is a geometric pattern matcher. It compares measured strain to predicted strain, and "predicted strain" is ultimately a statement about curvature dynamics. Numerical relativity enters because it calibrates and validates waveform families. When signals are loud, systematic waveform error becomes visible, and numerical geometry must be accurate enough that the remaining discrepancy is astrophysics, not numerics.

The geometry lesson is that scientific pressure scales with catalog size. As the catalog grows, we are forced to ask sharper questions: are there waveform systematics, are there subtle biases in spin inference, do ringdown modes agree with Kerr predictions, and how robust are our quasi-local diagnostics for the final black hole. Each sharper question reduces tolerance for numerical error.

This final figure is a practical translator. It tells you that a heavy system lives in the tens

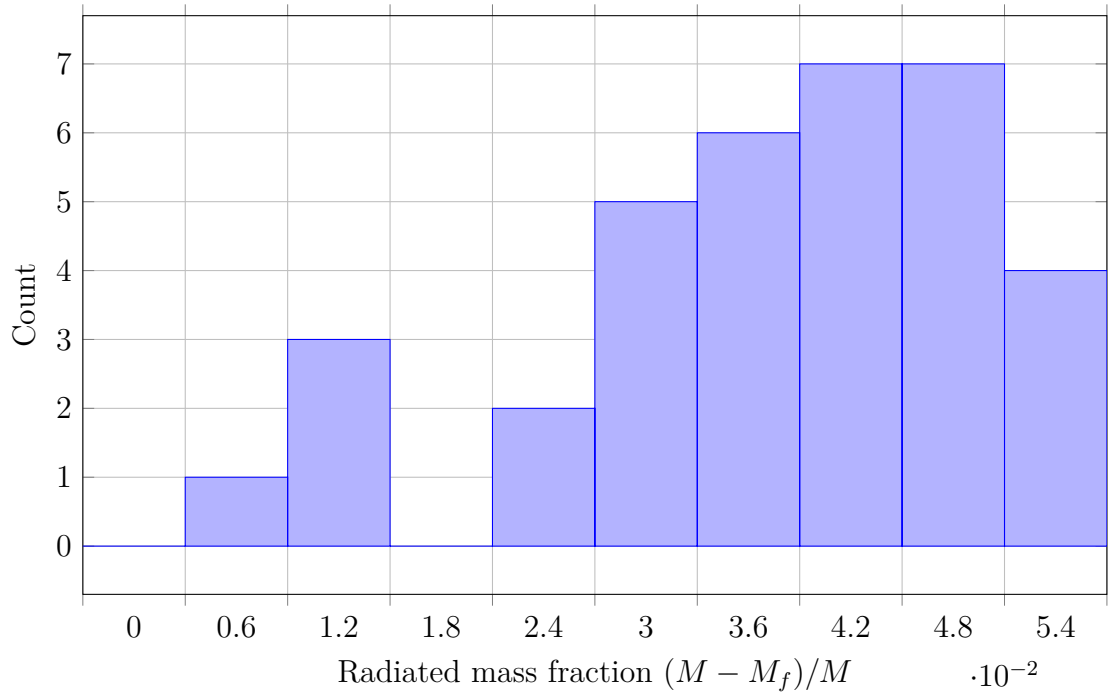


Figure 11: Data-driven Figure 23.6: Radiated mass fraction inferred from catalog medians via $f_{\mathrm{rad}} = (M - M_f)/M$ [3, 4]. This is a compact way to see how strongly nonlinear curvature dynamics convert mass into gravitational radiation, a core output that numerical relativity must capture.

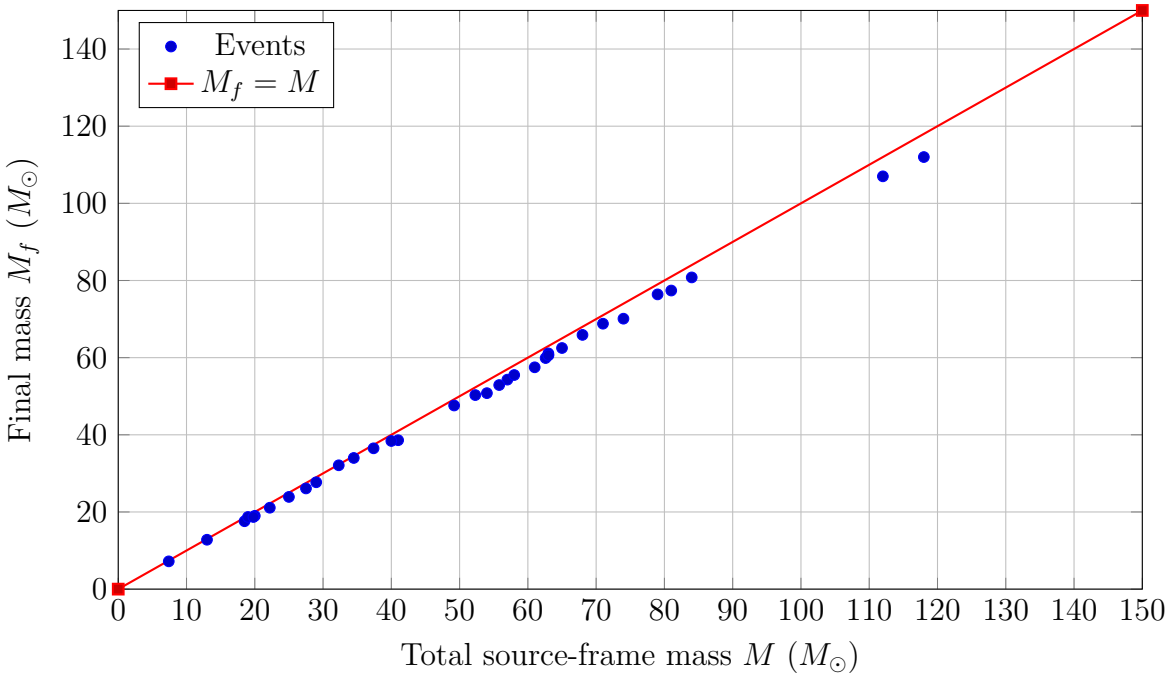


Figure 12: Data-driven Figure 23.7: Final mass versus initial total mass (source frame) from GWOSC default parameter estimates [3, 4]. Points below the diagonal reflect energy carried away by gravitational waves.

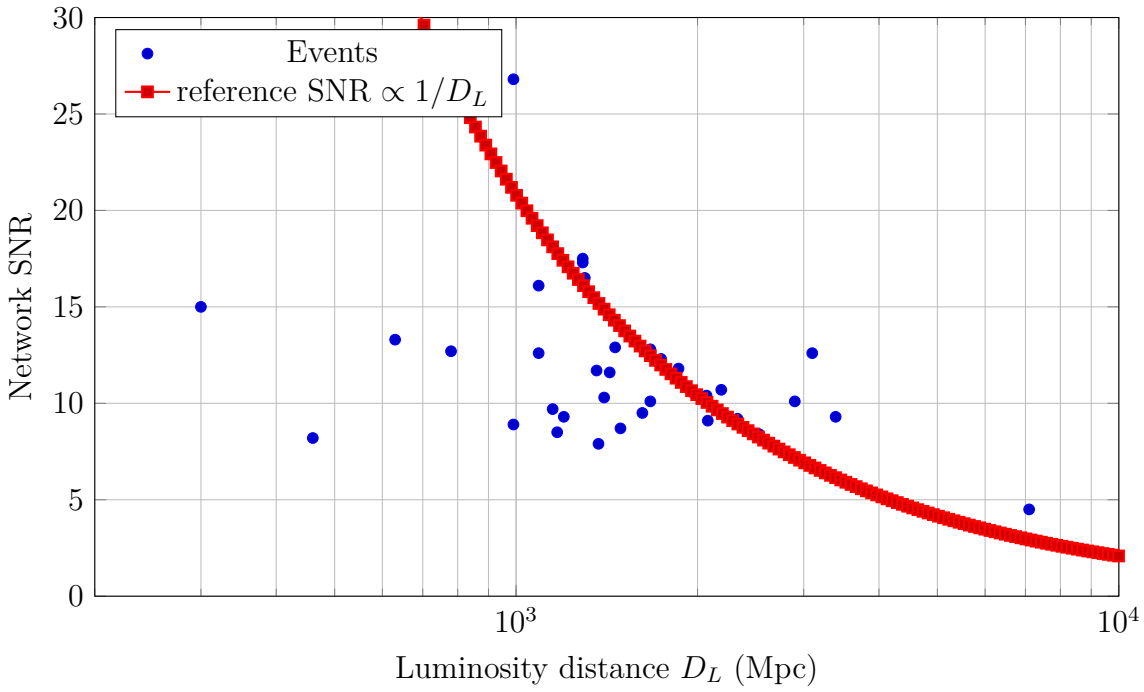


Figure 13: Data-driven Figure 23.8: Network matched-filter SNR versus luminosity distance for GWTC-3-confident events [3, 4]. A simple $1/D_L$ reference curve is plotted using a median scaling constant from the same dataset. The numerical relativity link is indirect but decisive: waveform fidelity controls matched filtering and therefore SNR recovery.

of Hz to low hundreds of Hz, where detectors are noisy and waveform systematics can matter in subtle ways. It also tells you that lighter systems move to higher frequencies where numerical resolution must handle faster phase evolution. In both regimes, the computational problem is geometric: represent curvature accurately across scales.

0.7 A workflow view: what a modern NR pipeline is really doing

0.8 Phenomenon statements: real-world applicability of computational differential geometry

This final section provides concrete phenomenon statements that connect the computational geometry of numerical relativity to observable or testable physics. Each statement is written as a geometric claim with an operational measurement attached.

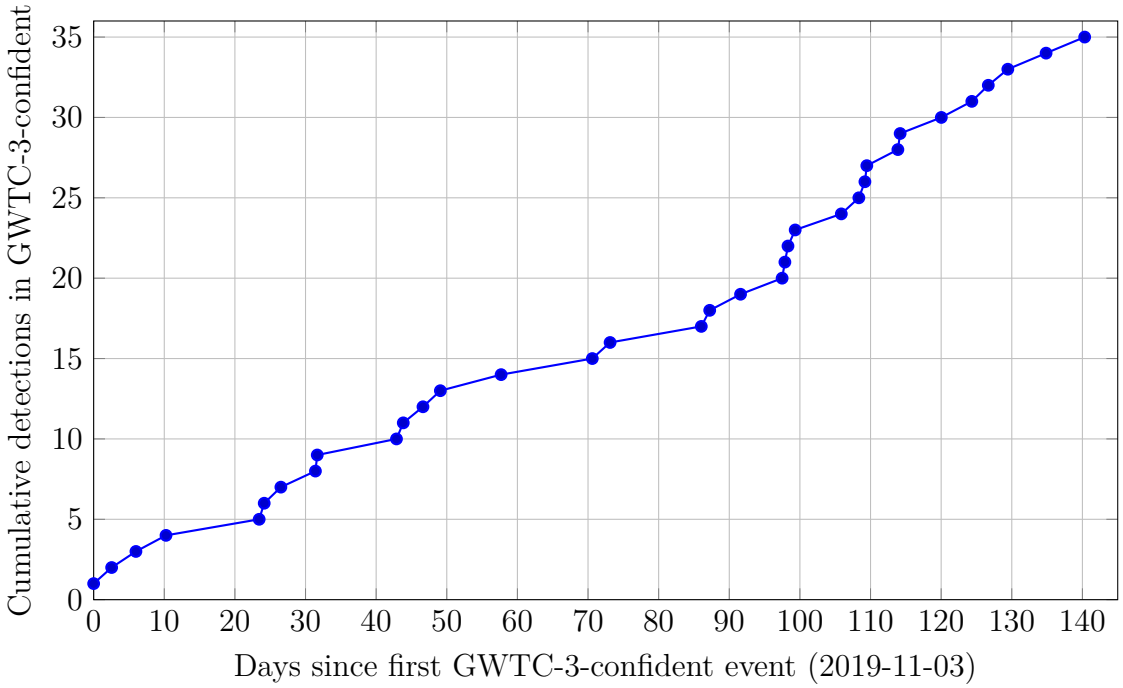


Figure 14: Data-driven Figure 23.9: Cumulative detections across the GWTC-3-confident time window, computed from event timestamps encoded in event names and sorted chronologically [2, 3]. Increasing event counts translate into increasing demands on waveform model coverage and numerical calibration.

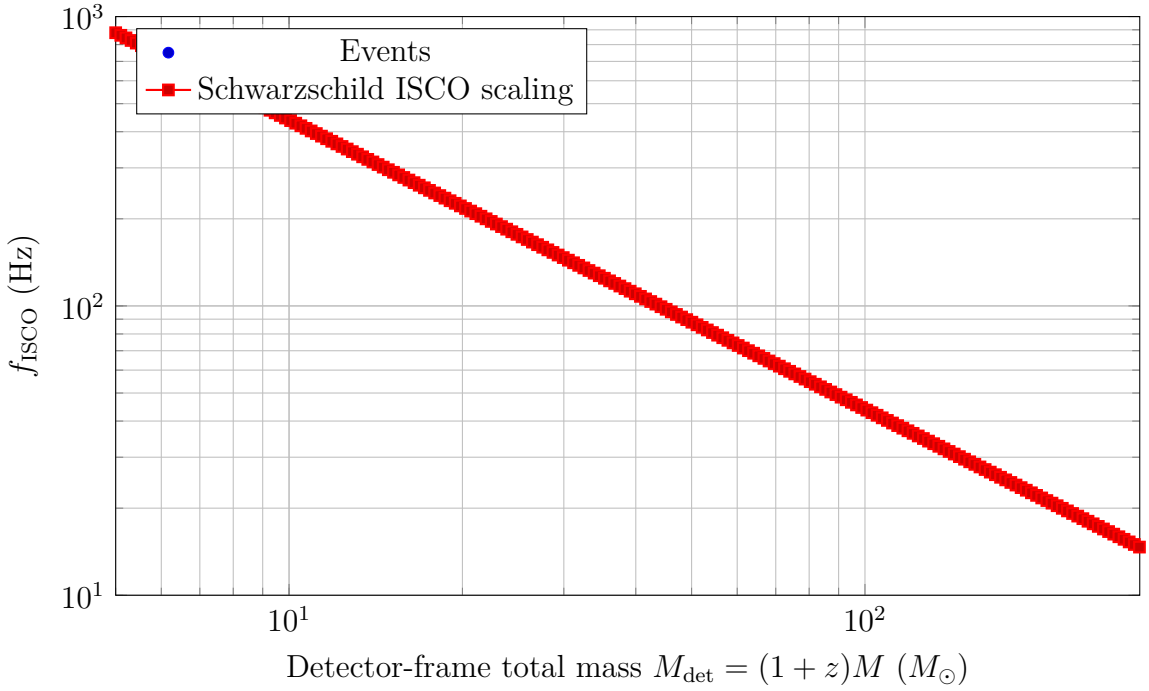
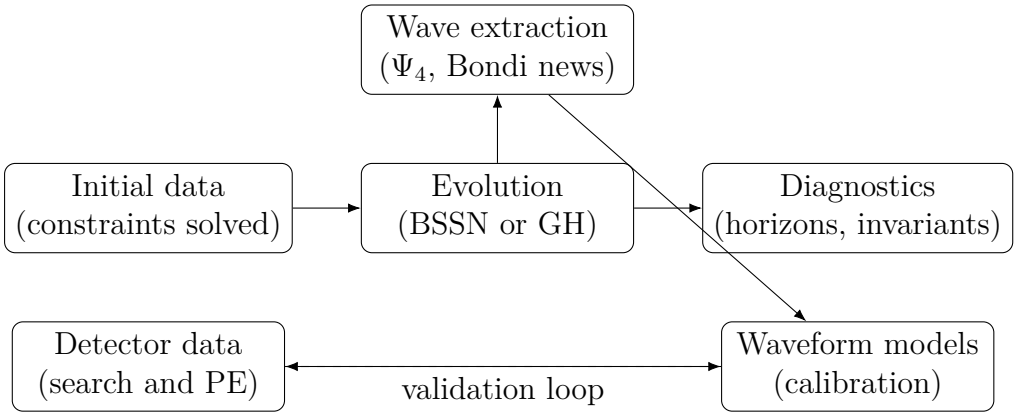


Figure 15: Data-driven Figure 23.10: An illustrative frequency scale for each event, computed from the detector-frame total mass using the Schwarzschild ISCO estimate (10) [3]. This plot is a bridge between geometry and instrument: masses map to frequencies, frequencies map to where numerical waveforms must be phase accurate.



Key idea: simulation and inference form a coupled loop.

Figure 16: Conceptual Diagram 23.F: A high-level workflow linking constraint-satisfying initial data to evolution, diagnostics, wave extraction, waveform modeling, and data analysis. The data-driven figures in this chapter use open catalog summaries as the observational boundary conditions for this loop [2, 4].

Phenomenon 1: Ringdown as a curvature spectroscopy experiment. When a binary black hole merger settles, the spacetime approaches Kerr and the remaining dynamics are quasinormal modes, which are eigen-solutions of the perturbed curvature. The practical implication is that numerical relativity must deliver a clean final mass and spin, plus a waveform whose late-time phase and amplitude are not polluted by gauge transients. In observational terms, the same requirement appears as the ability to test whether measured ringdown frequencies match Kerr predictions given the inferred remnant. Computational differential geometry contributes by providing quasi-local horizon diagnostics (mass, spin, multipoles) and by controlling wave extraction so that the asymptotic signal is accurate and tetrad-consistent [12, 13, 16].

Phenomenon 2: Unequal-mass binaries create multi-scale geometry. A large mass ratio binary is a geometric two-scale problem: one horizon is small, the other is large, and the curvature gradients near the small object can be extreme compared to the orbital scale. Numerically this forces refinement strategies, careful interpolation, and stable gauge drivers that do not introduce spurious coordinate shocks. Observationally, this regime matters because even a small object can imprint distinctive harmonics and precession signatures in the waveform. Figure 7 is therefore not merely a population plot; it is an empirical map of where a code must remain stable under strong scale separation [3].

Phenomenon 3: Energy radiated in gravitational waves is a global geometric flux. The mass loss $M - M_f$ is a covariant statement about energy carried to infinity, and it is tightly linked to the curvature content of the spacetime. Numerically, it demands a stable evolution plus a reliable computation of radiation at large radius or at I^+. Observationally, the same flux controls the relationship between inspiral masses and remnant properties used in data analysis. Figures 11 and 12 show that the radiated fraction is typically a few percent in GWTC-3-confident events, which is large enough that modeling errors at the percent level become scientifically meaningful [3, 4].

Phenomenon 4: Redshift mixes geometry with cosmology even when you do not mean to do cosmology. The detector sees a redshifted waveform. The source-frame evolution is what the binary "does", but the detector-frame masses are what control the observed time and frequency scales. Figure 10 makes this operational: it is a dataset-based reminder that any attempt to compare numerical waveforms to data must track $(1 + z)$ consistently. Inaccurate bookkeeping here is not a small clerical error; it is a phase error in the fundamental observable [2, 4].

Phenomenon 5: Wave extraction is a coordinate-patching problem with physical consequences. In the near zone, coordinates are chosen for stability and horizon handling. In the wave zone, coordinates should resemble inertial observers. This mismatch is why characteristic extraction and other methods exist: they are geometric patching strategies that attempt to compute radiation where it is cleanly defined. Observationally, this affects parameter inference and tests of general relativity that rely on small differences between models. Computationally, it affects how one validates that numerical errors are smaller than the science questions being asked [14, 15].

Phenomenon 6: Multi-messenger neutron star mergers expose the limits of pure vacuum geometry. Although this chapter emphasizes geometric infrastructure, neutron star mergers require coupling geometry to relativistic hydrodynamics and microphysics. The geometric core remains: constraints, gauge, curvature diagnostics, and extraction. The observational payoff is broader: gravitational waves plus electromagnetic signals can constrain the equation of state. In computational terms, the challenge is that matter introduces discontinuities and shocks, which interact with discretization choices. A geometry-aware discretization strategy must therefore coexist with shock-capturing methods, without sacrificing constraint control or horizon diagnostics when a black hole forms.

0.9 Conclusion

Numerical relativity is best understood as computational differential geometry: a consistent attempt to represent a Lorentzian manifold and its curvature in discrete form, while preserving the constraints and causal structure that make the continuum theory meaningful. The key lesson is not that one should memorize a particular formulation. The key lesson is diagnostic discipline. Every numerical result should be accompanied by a geometric interpretation (which object is this), a constraint story (why it is stable), and an extraction story (how it becomes an observable).

The ten dataset-based figures built from GWTC-3-confident event summaries highlight an empirical fact: the universe supplies binaries across a broad range of mass, mass ratio, spin, and redshift. That diversity is the real input specification for numerical relativity. If a method is unstable or inaccurate in one corner of parameter space, the catalog will eventually find it. The frontier is therefore not only faster computation, but more geometric robustness: better constraint control, cleaner radiation at infinity, more reliable quasi-local invariants, and end-to-end error budgets that survive contact with precision gravitational-wave astronomy.

Bibliography

[1] Gravitational Wave Open Science Center (GWOSC), "GWOSC API v2 Documentation," accessed Feb. 24, 2026. [Online]. Available: https://gwosc.org/api/v2/docs

[2] GWOSC, "All Events (Event Portal): default search and default parameter estimation values," accessed Feb. 24, 2026. [Online]. Available: https://gwosc.org/eventapi/html/allevents/

[3] GWOSC, "Event versions API endpoint (CSV with default parameters)," accessed Feb. 24, 2026. [Online]. Available: https://gwosc.org/api/v2/event-versions?format=csv&include-default-parameters=true&lastver=true&release=GWTC-1-confident%2CGWTC-2.1-confident%2CGWTC-3-confident%2CGWTC-4.0

[4] R. Abbott *et al.*, "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run," *Physical Review X*, vol. 13, 041039, 2023. [Online]. Available: https://journals.aps.org/prx/abstract/10.1103/PhysRevX.13.041039

[5] A. Trovato *et al.*, "GWOSC: The Gravitational Wave Open Science Center," *PoS Asterics2019*, 082, 2020. [Online]. Available: https://pos.sissa.it/357/082/

[6] M. Shibata and T. Nakamura, "Evolution of three-dimensional gravitational waves: Harmonic slicing case," *Phys. Rev. D*, vol. 52, pp. 5428-5444, 1995. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.52.5428

[7] T. W. Baumgarte and S. L. Shapiro, "On the numerical integration of Einstein's field equations," *Phys. Rev. D*, vol. 59, 024007, 1998. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.59.024007

[8] F. Pretorius, "Evolution of Binary Black-Hole Spacetimes," *Phys. Rev. Lett.*, vol. 95, 121101, 2005. [Online]. Available: https://arxiv.org/abs/gr-qc/0507014

[9] C. Gundlach, G. Calabrese, I. Hinder, and J. M. Martin-Garcia, "Constraint damping in the Z4 formulation and harmonic gauge," *Class. Quantum Grav.*, vol. 22, pp. 3767-3774, 2005. [Online]. Available: https://arxiv.org/abs/gr-qc/0504114

[10] L. Lindblom, M. A. Scheel, L. E. Kidder, R. Owen, and O. Rinne, "A New Generalized Harmonic Evolution System," 2005. [Online]. Available: https://arxiv.org/abs/gr-qc/0512093

[11] F. Loffler *et al.*, "The Einstein Toolkit: a community computational infrastructure for relativistic astrophysics," *Class. Quantum Grav.*, vol. 29, 115001, 2012. [Online]. Available: https://doi.org/10.1088/0264-9381/29/11/115001

[12] J. Thornburg, "Event and apparent horizon finders for 3+1 numerical relativity," *Living Reviews in Relativity*, vol. 10, article 3, 2007. [Online]. Available: https://link.springer.com/content/pdf/10.12942/lrr-2007-3.pdf

[13] E. T. Newman and R. Penrose, "An Approach to Gravitational Radiation by a Method of Spin Coefficients," *Journal of Mathematical Physics*, vol. 3, pp. 566-578, 1962. [Online]. Available: https://pubs.aip.org/aip/jmp/article/3/3/566/387969/An-Approach-to-Gravitational-Radiation-by-a-Method

[14] N. T. Bishop *et al.*, "Cauchy-characteristic extraction in numerical relativity," 1997. [Online]. Available: https://arxiv.org/abs/gr-qc/9705033

[15] C. Reisswig, N. T. Bishop, D. Pollney, and B. Szilagyi, "Characteristic extraction in numerical relativity: binary black hole merger waveforms at null infinity," 2010. [Online]. Available: https://arxiv.org/abs/0912.1285

[16] J. L. Ripley *et al.*, "Gravitational-Wave Parameter Inference with the Newman-Penrose Scalar," *Physical Review X*, vol. 13, 041048, 2023. [Online]. Available: https://journals.aps.org/prx/abstract/10.1103/PhysRevX.13.041048

[17] M. J. Berger and J. Oliger, "Adaptive mesh refinement for hyperbolic partial differential equations," *J. Comput. Phys.*, vol. 53, pp. 484-512, 1984. [Online]. Available: https://doi.org/10.1016/0021-9991(84)90073-1

Index

Ψ_4, vii
3+1 decomposition, iii

adaptive mesh refinement, v

BSSN, vi

Cauchy-characteristic extraction, viii

extrinsic curvature, iv

generalized harmonic gauge, vii

Newman-Penrose, vii
numerical relativity, ii